Impact Assessment of Soyabean Grain Storage Facilities on Deforestation Detected through Remote Sensing Images -The Brazil Case Study.

CHARLIE SHAWA
Enschede, The Netherlands, July, 2024

SUPERVISORS:
Dr. Yue Dou
Dr. Claudia Paris

Impact Assessment of Soyabean Grain Storage Facilities on Land cover and Deforestation Detected through Remote Sensing Images - The Brazil Case Study.

CHARLIE SHAWA
Enschede, The Netherlands, July, 2024

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Resources Management

SUPERVISORS:
Dr. Yue Dou
Dr. Claudia Paris
THESIS ASSESSMENT BOARD:
Prof.dr.ir. L.L.J.M. Willemen (Wieteke) (Chair)
Dr Ramon Felipe Bicudo da Silva (External Examiner, State University of Campinas, Brazil)

Abstract

The expansion of agricultural land is one of the major contributors to deforestation, biodiversity loss, and greenhouse gas emissions particularly in tropical regions. Soyabean cultivatable land especially in the tropical regions of South America has increased significantly. This is driven by rising global population and meat consumption which boosts the demand for soyabean-based animal feed. This expansion has subsequently led to the growth of agricultural logistical infrastructure such as grain storage facilities which play a pivotal role in its supply chain. However, the role of such logistical infrastructure in relation to deforestation is unknown.

To address this knowledge gap, this study introduces a novel spatial-temporal analysis method using buffer zones around grain storage facilities established between 2002 and 2017 in Mato Grosso. Buffer zones of 5 km, 10 km, and 25 km around storage facilities were used as spatial extents within which land cover changes (deforestation) were analysed. Using geoinformation science techniques and statistical models namely propensity score matching and logistic regression, the study explored deforestation dynamics as well as the associated effects of soyabean grain facilities (treatment), biophysical, socioeconomic and climates factors.

The study findings reveal that deforestation rates in the buffer zones significantly decreased between 2003 and 2008 before stabilizing at lower levels. Despite the localized approach using buffers, this aligns with the observed peaks and troughs in deforestation trends at broader national and international resulting from policies such as the soy moratorium and the Action Plan for Deforestation Prevention and Control in the Legal Amazon (PPCDAm). The paired t-test shows substantial deforestation reductions near soyabean facilities, highlighting the effectiveness of close-proximity interventions, likely due to better monitoring, enforcement, and adoption of sustainable practices. Conversely, the odds ratios indicate that slope, GDP per capita, and the human influence index are associated with lower deforestation odds. However, soyabean facilities increase the overall odds of deforestation. This paradox occurs because while soyabean facilities drive localized reductions in deforestation due to effective interventions, they also promote broader agricultural expansion, infrastructure development, and economic incentives for land conversion, leading to increased deforestation beyond immediate proximities.

These findings emphasize the importance of integrating economic development with environmental sustainability to mitigate deforestation effectively in the municipalities of Mato Grosso. This in turn should inform policies that balance agricultural expansion and environmental conservation.

Acknowledgements

I would like to express my deepest gratitude to all those who have supported and contributed to the successful completion of this research work.

First and foremost, I extend my sincere thanks to my supervisors Dr. Y. Dou and Dr. C. Paris for their invaluable guidance, constructive feedback and constant encouragement throughout this research journey. Their expertise and insights have been instrumental in shaping the direction and outcome of this study. I would like to express my deep appreciation to Dr. Raymond G. Nijmeijer, the NRM Program Director for his supportive and considerate approach during my two years of master's studies particularly during the thesis phase. Furthermore, I am thankful to Prof.dr.ir. L.L.J.M. Willemen (Wieteke), the Chair of this research work for her valuable insights and constructive feedback. I am grateful to the external examiner, Dr Ramon Felipe Bicudo da Silva, for investing your time and providing profound insights during the final defense.

Special thanks to the Faculty of Geo-information Science and Earth Observation (ITC) community at the University of Twente for providing a conducive and stimulating environment for research. The resources, facilities, and academic atmosphere at have been fundamental in facilitating this work.

I would also like to extend my pofound gratitude to the Dutch government and the Orange Knowledge Programme (OKP) for their generous financial support which has been essential in making my study and indeed this research possible.

I am profoundly grateful to my friends, my support system, who have been instrumental in my journey. Mwangala Simate, my sister and countrymate, you made me feel at home away from home. Anesu Chimbi, Calvin Swai and Tionge Munthali, my colleagues from Southern Africa, you were like family during my stay and study in the Netherlands. To my colleagues from Western Africa, Abdul Cissey, Lawrence Lemgo, Prince Boateng, and Emmanuel Salami, your camaraderie and lively spirit have made my experience unforgettable. Thank you for your unwavering support and motivation. Your friendship has been a cornerstone of my journey.

To my fiancé, Precious Tembo I am eternally grateful for your understanding, patience and encouragement. Your support has been a cornerstone of my perseverance and success. Your belief in me has been a constant source of strength, and I am deeply grateful for your presence in my life.

A special thanks family whose belief in me has been a driving force throughout my academic endeavours. Your sacrifices and constant support have been deeply appreciated.

TABLE OF CONTENTS

List of figures

List of tables

List of abbreviations

CAR	Cadastro Ambiental Rural, or Rural Environmental Registry
DEM	Digital elevation model
DR	Deforestation rate
GDP	Gross domestic product
HII	Human influence index
PPCDAm	Portuguese acronym for Action Plan for Deforestation Prevention and Control in the Legal Amazon
SMD	Standardized Mean Difference

Chapter One: Introduction

1.1. Background

Global land cover change and soyabean production

Globally, the effects of human-induced environmental transformations such as biodiversity loss and deforestation cannot be overemphasized. Álvaro et al., (2012) define biodiversity loss as the decline in the variety of life forms within an ecosystem including species extinction, genetic diversity reduction and ecosystem degradation while deforestation as the large-scale removal of forest cover often for agricultural or urban development purposes. Between 1960 and 2019, the extent of the effects of these human-induced environmental transformations was estimated to be about four times greater than previously calculated from long-term land change assessments (Winkler et al., 2021).

A significant driver of both biodiversity loss and deforestation particularly in the tropical regions is the expansion of agricultural land (López-Carr, 2021). The growing global demand for commodities such as oil palm and soyabeans has intensified this issue fuelling critical debates on balancing agricultural expansion with forest conservation and ecosystem services preservation (Macedo et al., 2012). Specifically, the production area for soyabeans is projected to increase due to rising global population and meat consumption which boosts the demand for soy-based animal feed (Masuda & Goldsmith, 2009; Vivek et al., 2020). Between 1968 and 2018, soyabean production increased to about 350 million tonnes which has been harvested over almost 125 million hectares, with the United States of America (USA), Brazil, Argentina, China, and India accounting for over 85% of global production (De Maria et al., 2020; Shaik & Kwame-Asiam, 2020).

Besides the economic benefits associated with soyabean in Brazil, its expanding production has been linked to deforestation, land degradation and greenhouse gas emissions (Fehlenberg et al., 2017). A study by Song et al., (2021) highlights the significance of deforestation driven by soyabean expansion in the Brazilian Amazon and Cerrado noting a 160% increase in harvested soyabean area in Brazil since 2000. However, a study by Filassi & de Oliveira, (2021) notes that despite Brazil being one major producer of soyabean, it faces a significant shortage of storage facilities for agricultural products. The study reveals the deficit in storage capacity has grown from 6.6 million tons in the 2008-09 crop year to 76.0 million tons in the 2018-19 crop year. This lack of storage infrastructure affects producers the most as they often must sell their produce immediately, leaving them little bargaining power to the price and transferring the storage problem to the next agents in the supply chain.

The shortage of storage facilities not only creates economic inefficiencies but also worsens environmental issues. Without adequate storage, the pressure to clear more land for immediate cultivation increases, thereby contributing to further deforestation (Szerman et al., 2022). However, the extent of deforestation and land cover dynamics related to soyabean expansion and associated infrastructure, particularly grain storage facilities remain unclear. Addressing the associated effects of such agriculture infrastructure on deforestation crucial for meeting climate targets outlined in Sustainable Development Goals 13 and 15 (The Sustainable Development Goals, 2015). A key objective in mitigating climate change and biodiversity loss is achieving zero deforestation (Erasmus et al., 2020).

Grain storage facilities and soyabean storage in Brazil

The expanding soyabean production in Brazil has put pressure on agriculture grain storage facilities critical for post-harvest loss management (Fernando et al., 2016). Despite annual increases in grain output, the number of grain storage facilities has not changed considerably (Eberhardt et al., 2017; Erasmus et al., 2020). Grain storage facilities as critical nodes for food security play a crucial role in the spatial distribution and cultivation of soyabean and other crops. Nonetheless, a study by Del et al., (2009) underscores the environmental impacts of soyabean expansion driven by infrastructure development such as roads and storage highlighting biodiversity loss due to agricultural land expansion. A study by Garrett et al., (2013) emphasizes the influence of supply chain configurations on soyabean production highlighting that higher yields and planted areas are in regions with improved infrastructure. The study further reveals that the global factors driving soyabean expansion include technological changes facilitated by efficient storage and transportation systems.

It can be concluded from these studies that grain storage facilities are key infrastructural components which influence the spatial distribution of cultivatable lands consequently deforestation. However, they have not addressed the spatial-temporal impact associated with grain storage facilities on land cover changes and deforestation. Assessing the long-term effect of grain storage facilities on deforestation requires addressing the complex relationship between these facilities, land cover change and other socio-ecological drivers (Gladek et al., 2017). This study aims to address this gap by using buffer zones of varying sizes as observational framework for the multitemporal assessment of how grain storage facilities influence deforestation and land use patterns at different spatial extents.

Buffer zones are defined as areas managed to enhance positive conservation impacts and reduce negative effects on neighbouring communities (Ebregt & De Greve, 2000; Perelló et al., 2012). According to the Brazilian National Environment Council, buffer zones are vital for regulating and minimizing negative human activities in the surroundings of conservation units. The Brazilian National Environment Council initially proposed a 10 km radius around conservation units. But this buffer zone size was later reduced to 3 km due to concerns raised by the productive sector (Bellón et al., 2020). However, in this study, buffers typically used for assessment on environmental impacts of agricultural production facilities have been considered. This ensures relevance and consistency in the analysis of grain storage facilities' impacts on deforestation and land use patterns over different spatial and temporal scale. Specifically, buffer zones of 5 km, 10 km, and 25 km will be utilized to assess the spatiotemporal effects of grain storage facilities on deforestation. The 5 km buffer is chosen to evaluate impacts in areas most susceptible to direct influence from the facilities. This is consistent with the findings of Del et al., (2009) which highlight the environmental impacts of soyabean expansion driven by infrastructure development in its vicinity.

Expanding the analysis to 10 km buffer allows for an examination of moderate-range effects encompassing changes in local infrastructure and ecological interactions influenced by the facility. This aligns with the need to understand a broader spectrum of impacts beyond immediate proximity. Lastly, the 25 km buffer will be used to provide insight into broader regional impacts including larger landscape-level changes and shifts in ecological patterns and socio-economic dynamics. Garrett et al., (2013) underscores the importance of considering supply chain infrastructure configurations on soyabean production at larger spatial scale particularly in regions with improved infrastructure.

The study will examine the spatial-temporal impact associated with grain storage facilities on land cover changes and deforestation in Brazil's state of Mato Grosso, a major soyabean producer. The considered timeframe, spanning from 2002 to 2017, covers different phases of deforestation, policy actions and agricultural growth in the state. According to Kazadi & Yoshikawa, (2009), during the period between 2001 and 2004, the state experienced high deforestation rates driven by agricultural expansion especially soyabean cultivation and favourable market conditions. Between 2005 and 2017, there were fluctuating deforestation rates influenced by policy interventions such as the Amazon Soyabean Moratorium and the Cattle

Agreements as well as market forces, climate variability and fire events. Furthermore, the state stands out due to its geographic diversity and economic significance. It is covered by the Cerrado, Amazon and Pantanal biomes and has undergone significant land cover transformations with large, forested areas being converted to agricultural lands primarily for soyabean and beef production since the beginning of the century (Kuschnig et al., 2021).

Methodological approaches

A study by Singh et al., (2024) highlights two fundamental steps in any study of land use land cover change detection namely identifying changes in the landscape and determining the causes of those changes. Furthermore, advances in the acquisition, processing and interpretation of remotely sensed imagery have significantly facilitated the detection of changes in land cover. However, the study notes that explaining these changes by identifying and attributing them to specific factors and their causal effects remains challenging.

Various studies have employed remote sensing, geoinformation science, and statistical models to understand the drivers of deforestation. Empirical statistics, particularly logistic regression models play a crucial role in understanding how variables namely socio-economic and biophysical variables known to shape humans and forests interactions e.g., agricultural expansion influence deforestation (Yang et al., 2021). These models systematically test assumptions such as the relationship between variables being linear in the logit form and the absence of multicollinearity among predictors. They also help rank the importance of variables and provide insights into the underlying processes of deforestation (Müller et al., 2011; Rosa et al., 2013).

This study will explore how the socio-economic and biophysical variables influence deforestation by focusing on defined zones of grain storage facilities in the state of Mato Grosso, Brazil. A matching method will be employed to evaluate the impact of these variables. Matching is a statistical technique used to compare treated and control groups that are similar in covariates(Austin, 2011). It provides a robust evaluation with fewer assumptions about the underlying data distributions and relationships (Garrido et al., 2014). This approach differs from classical experimentation as it involves quasi-experimental designs like natural experiments. These experiments identify real-world situations to control for as many potential causal factors as possible while focusing on variation in one key factor (Busch & Ferretti-Gallon, 2017). The matching method will help underscore the association of grain storage facilities on deforestation considering also the underlaying socioeconomic and biophysical variables. In addition to matching, logistic regression will be used to further analyse and quantify the relationships between these variables and deforestation outcomes. Logistic regression allows for the examination of how various predictors (both categorical and continuous) influence the likelihood of deforestation occurring in the vicinity of grain storage facilities. This statistical approach will provide complementary insights into the factors driving land use changes in the study area.

Together, these methods will enable a comprehensive assessment of the socio-economic and biophysical determinants of deforestation associated with grain storage facilities in Mato Grosso, Brazil, offering insights that can inform more effective conservation and land management strategies.

1.2. Objectives, research questions and hypothesis

Main objective

The main objective of this research is to investigate to what degree soyabean grain storage facilities contribute to deforestation in Brazil considering their distribution over space and time.

Specific objective 1

1. To assess whether land cover change before and after can be associated with the appearance of soyabean grain storage facilities in the municipalities of Mato Grosso from 2002 to 2017 within 5, 10 and 25km buffers.

Research questions.

a. How does land cover change around Mato Grosso's soyabean facilities vary spatially from 2002 to 2017?
b. What are the spatial patterns of land cover change in buffer zones surrounding soyabean facilities in Mato Grosso from 2002 to 2017?

Specific objective 2

2. To investigate the impact associated with soyabean grain storage facility on deforestation between 2002 and 2017.

Research questions.

a. What are the effects associated with soyabean grain storage facility on deforestation considering the impact of socioeconomic, biophysical and climate factors?

Hypothesis

The study hypotheses the following for objective one:

a. **Null hypothesis**: The establishment of soyabean grain storage facilities in Mato Grosso between 2002 and 2017 is not associated with significant changes in land cover within defined buffer zones around the facilities.
b. **Alternative hypothesis:** The establishment of soyabean grain storage facilities in Mato Grosso between 2002 and 2017 is associated with significant changes in land cover within defined buffer zones around the facilities.

The study hypotheses the following for objective 2:

a. **Null hypothesis:** Soyabean grain storage facilities did not contribute to deforestation in Mato Grosso, Brazil between 2002 and 2017.
b. **Alternative hypothesis:** Soyabean grain storage facilities contributed to deforestation in Mato Grosso, Brazil between 2002 and 2017.

Chapter Two: Methodology

2.1. Study area

Mato Grosso, situated in central Brazil, stands out for its distinct natural features and significant economic contributions to the country's agricultural sector. The state is considered a key agricultural region in Brazil due to its extensive arable land and favourable soil conditions. The state plays a significant role in national agriculture, particularly in soyabean production. The state spans approximately 903,357 km² and located in the middle of South America (7° to 18°S and 50° to 61°W). It is intersected by three major Brazilian biomes: the Amazon, covering 54% in the north and parts of the west; the Cerrado, encompasses 39.5% of the state's centre; and the Pantanal, comprising 6.5% in the southwest (Brandão et al., 2019).

Mato Grosso has a history of substantial deforestation and was once considered a global hotspot notably in the early 2000s (DeFries et al., 2013). This deforestation, driven by activities like cattle ranching and soyabean production for distant markets, has witnessed a remarkable decline over time. Factors contributing to this decline include changes in market dynamics, policy implementations, enhanced enforcement measures, and improved monitoring systems (Kuschnig et al., 2021). The transformation from a deforestation hotspot to a region with declining rates underscores Mato Grosso's significance as a case study area. Studying its journey provides valuable insights into deforestation drivers, the effectiveness of conservation policies, and the potential for replicating successful strategies globally.

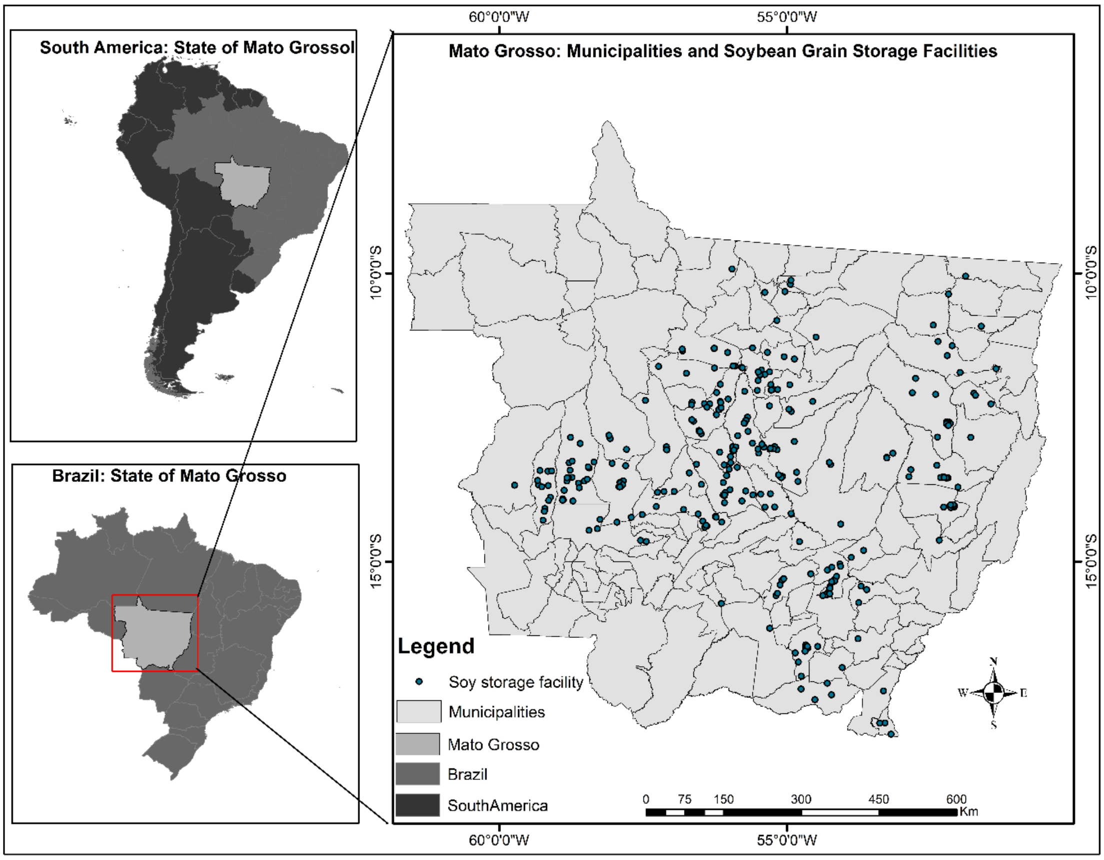


Figure 1: A map showing the location of Brazil in South America, Mato Grosso in Brazil and the distribution of soyabean storage facilities in 2017.

.

2.2. Study dataset and sources

Numerous studies have been undertaken on the variables shaping land cover changes and deforestation with notable contributions from researchers such as Garrett et al. (2013), Meier et al., (2021), Opršal et al. (2016) and Samuel (2011). The table below outlines the relevant data sources and variables.

Table 1: Socio-economic and biophysical variables of the study region

Data	Description and Units of measure	Spatial resolution	Temporal coverage	Data source
Land cover maps	Land cover type map (Mapbiomas 7.1 thematic maps)	30m	2002-2017	https://brasil.mapbiomas.org/en/colecoes-mapbiomas/
Digital elevation model (DEM)	Shuttle Radar Topographic Mission (SRTM)/Meter	90m	2018	https://srtm.csi.cgiar.org/download
Soil	Soil clay/ Percentage	250m	2005	https://files.isric.org/soilgrids/former/2017-03-10/data/
	Soil pH			
	Soil depth/Centimetres			
	Soil bulky density/ Kilogram per cubic meter			
	Organic Carbon Stock/Ton per Hectare			
Climate	Precipitation/ mm	5.55km	2018	https://data.chc.ucsb.edu/products/CHIRPS-2.0/
Soyabean grain storage facility (Silo)	Metallic storage facilities		2002-2017	https://supplychains.trase.earth/logistics-map.
Global Accessibility Map	Travel time to a location of interest-based travel.	1km²	2008	https://forobs.jrc.ec.europa.eu/gam/download.
Population	Number of people/Million	Persons Per Municipality	2002-2017	https:// sidra. ibge. gov. br/ pesqu isa/ estim apop/ tabel as Sistema IBGE de Recuperação Automática - SIDRA
Gross Domestic Product	Market value of a country's goods and services in a time period/Real in Million	Reais per Million	2002-2017	https://sidra.ibge.gov.br/pesquisa/pib-munic/tabelas).
Human Influence Index	Human impact measure based on land use, population, and other factors/Dimensionless	-	2005	Global Human Influence Index (Geographic), v2: Last of the Wild, v2 | SEDAC (columbia.edu)
Admin Boundary Data	Administration 0,1 and 2 boundaries	-	2020	https://data.humdata.org/dataset/cod-ab-bra?

2.3. Data preparation and application

The data preparation process involved several key steps to ensure the datasets were suitable for analysis:

Land cover data.

The downloaded Mapbiomas thematic maps based on Landsat satellite imagery from 1985 to 2021 provided detailed land cover information with a 30-meter spatial resolution and an accuracy of 88.1%. These maps classify diverse land cover categories, such as natural forests, urban infrastructure, water bodies, and agricultural areas. These maps were used to create annual land cover change maps used in the analysis.
Before analysis, the thematic maps were reprojected to WGS_1984_Web_Mercator projection. Annual land cover change maps were created from these maps using pixel-level change detection analysis through pairwise comparisons. The creation of the annual land cover change maps involved creation of a binary forest mask for each year distinguishing forested (value of 0) from deforested areas (value of 1). These masks were multiplied with the land cover map of the following year allowing for tracking changes in forest cover. Through pixel-wise multiplication, areas identified as forested in the previous year's mask were retained while non-forest areas were masked out. This method produced annual change maps that highlight shifts in forest cover. These maps are vital for identifying regions where forested land transitioned to other land cover types. The following figure illustrates the pixel-wise multiplication process.

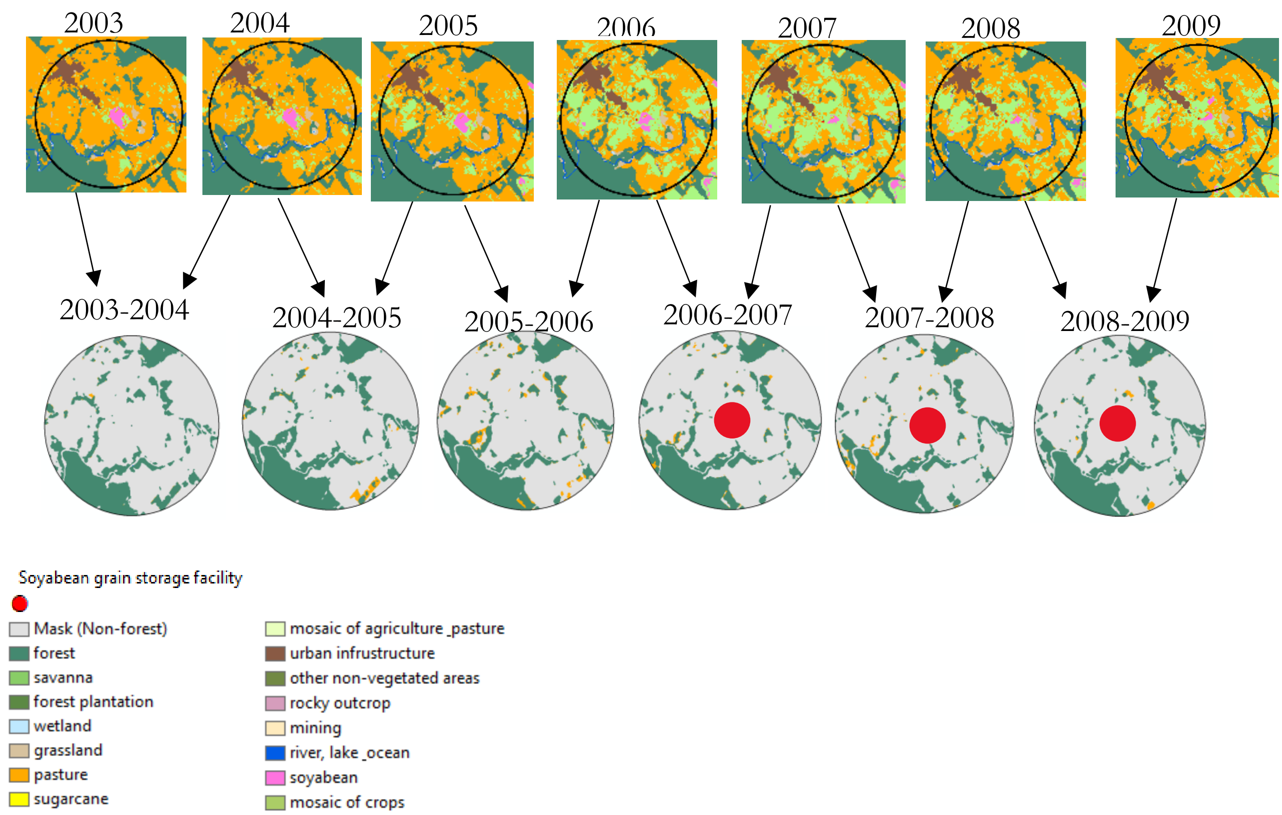


Figure 2: Demonstration of annual land cover change map creation.

Topological variables

Topological variables namely aspect, slope, and elevation were derived from the digital elevation model (DEM). These variables were clipped to the study area extent. They were then resampled to 30 meters using bilinear interpolation which calculates the weighted average of the four nearest input cells to match the spatial resolution of the land cover data. The datasets were then projected to WGS_1984_Web_Mercator projection. These operations were done in ArcMap.

Soil properties

Soil properties, including soil pH, percentage clay, bulk density, and soil carbon content, were processed similarly. These datasets were clipped to the study area extent, resampled to 30 meters using bilinear interpolation except for soil pH which was resampled using the nearest neighbourhood, and re-projected to the WGS_1984_Web_Mercator projection in ArcMap to ensure consistency with the land cover maps.

Precipitation data

Precipitation data were processed similarly. The data were clipped to the study area extent, resampled to 30 meters using bilinear interpolation, and projected to the WGS_1984_Web_Mercator projection to match the spatial resolution of other datasets.

Socioeconomic datasets

Socioeconomic datasets such as soyabean grain storage facilities, gross domestic product (GDP) and population data in excel format were filtered to match the study's timeframe (2002 to 2017). The soyabean grain facility data was converted to vector format (point) for further analysis. Using the GDP and population datasets, the GDP per capita was calculated by dividing GDP by population for each year then computing the average GDP per capita. The CSV file containing GDP per capita data for each municipality was joined to the corresponding municipalities in the shapefile. This joined shapefile data was then converted into a raster format. This ensured that the GDP per capita data was in spatial format suitable for analysis.

The temporal analysis of soyabean grain storage facilities from 2002 to 2017 was validated using Google Earth Pro. This verification exercise confirmed whether the soyabean grain facilities existing now also existed in previous years within the timeframe of 2002 to 2017.

Other socioeconomic variables, such as travel time and the human influence index (HII) were also clipped to the study boundary, resampled using nearest neighbourhood to a 30-meter resolution and projected to the WGS_1984_Web_Mercator projection.

2.4. Methods

Below is a description of the approaches adopted to assess the spatial-temporal impact of soyabean grain storage facilities on deforestation. Each step in the process is designed to contribute to achieving the first goal of the research:

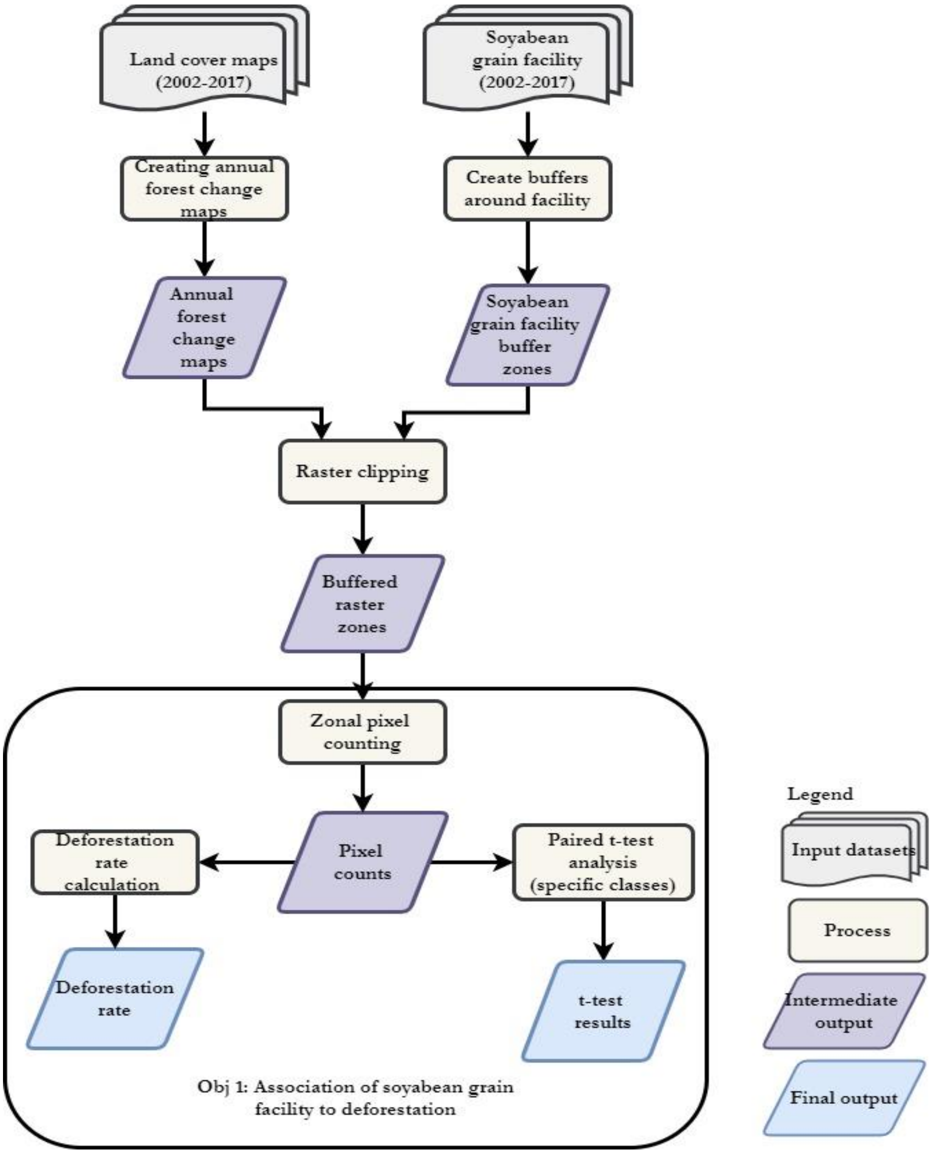


Figure 3: Methodology workflow for the first objective.

2.4.1. Association of land cover changes with soyabean facilities in Mato Grosso

a. Mean annual deforestation rates within buffers in Mato Grosso from 2002 to 2017

A spatial-temporal analysis approach was utilized to evaluate land cover changes surrounding grain storage facilities in Mato Grosso between 2002 and 2017. Annual deforestation rates (DR) were computed within buffer zones of 5, 10, and 25 kilometres around each facility to capture varying extents of influence. The DR computation involved taking the absolute of the comparison of forested areas between consecutive years within each buffer zone. This resulted in unique deforestation rate calculations for each buffer radius and facility location. The figure below demonstrates change maps from which the forest area was extracted:

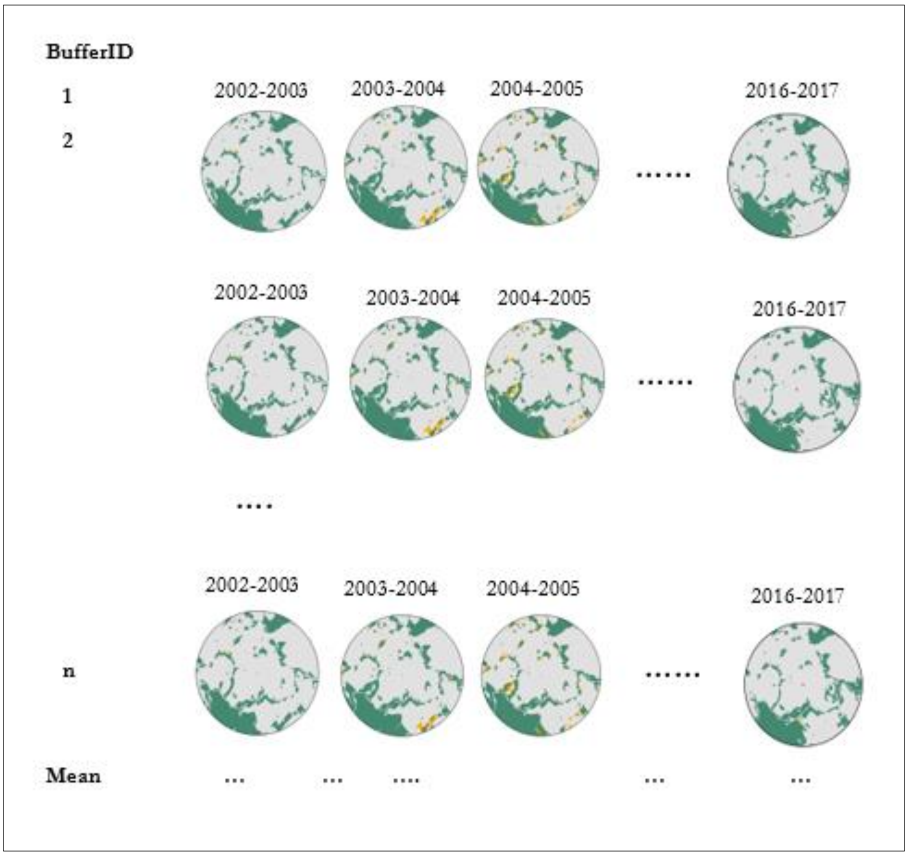


Figure 4: Annual change maps extracted to 5km buffer extent showing forested and deforested areas.

Higher DR values indicate faster forest loss, while lower rates suggest slower deforestation. By analysing deforestation rates across multiple buffer radii and facility locations, spatial-temporal trends in land cover changes associated with grain storage facilities were discerned. Mathematically, the formula for the annual deforestation rate (*DR*) for each year (*t*) is given by:

$$Deforestation\ rate\ (DR) = abs\left[\frac{(\omega^{y_t}-\omega^{y_{t-1}})}{\omega^{y_{t-1}}}\right]*100$$

Where:

- ω^{y_t} : represents the forest area present in the tth year.
- $\omega^{y_{t-1}}$: represents the forest area present in the (t-1)th year.

To compute the mean deforestation rate (MDR) per year, the annual deforestation rate was averaged by the number of buffers. The formular was as below:

$$\mathrm{MDR} = \frac{DR_t}{n_i}$$

Where n_i is the number of buffers, DR_t is the annual deforestation rate (*DR*) for each year (*t*).

b. Spatial impact assessment of soyabean facilities on deforestation rates and land cover change.

To assess whether land cover changes before and after the establishment of soyabean grain storage facilities in the municipalities of Mato Grosso from 2002 to 2017 are associated with these facilities, a paired t-test was used. The paired sample t-test is a widely used statistical method for determining differences between two measurements of a variable taken at two distinct times or under two related conditions (Kim et al., 2018; Xu et al., 2017).

In this study, changes in forest area were analyzed three years before and after the construction of the soyabean grain storage facility. The forest area that transitioned to other land cover classes was extracted from the generated annual change land cover maps using buffers of 5, 10, and 25 km as extents and the area was calculated.

The deforested area across buffers for the three years before and three years after the establishment of the soyabean grain storage facility were derived and used as input variables in the paired t-test. The selection of a three-year temporal window aligns with ecological and agricultural understanding of land cover dynamics. Significant changes in land cover particularly in response to anthropogenic interventions like the construction of soyabean grain facilities often take atleast three years to show (Roy et al., 2022).(Roy et al., 2022). This timeframe allows for capturing both immediate and slightly delayed effects thereby reducing short-term variability and minimizing the influence of long-term changes.

Below is a figure demonstrating at pixel-level the process of deriving deforested areas. The figure shows that before the soyabean grain facility, there was one change (forest patch to another class), and after the soyabean grain facility was established, the changes increased to three.

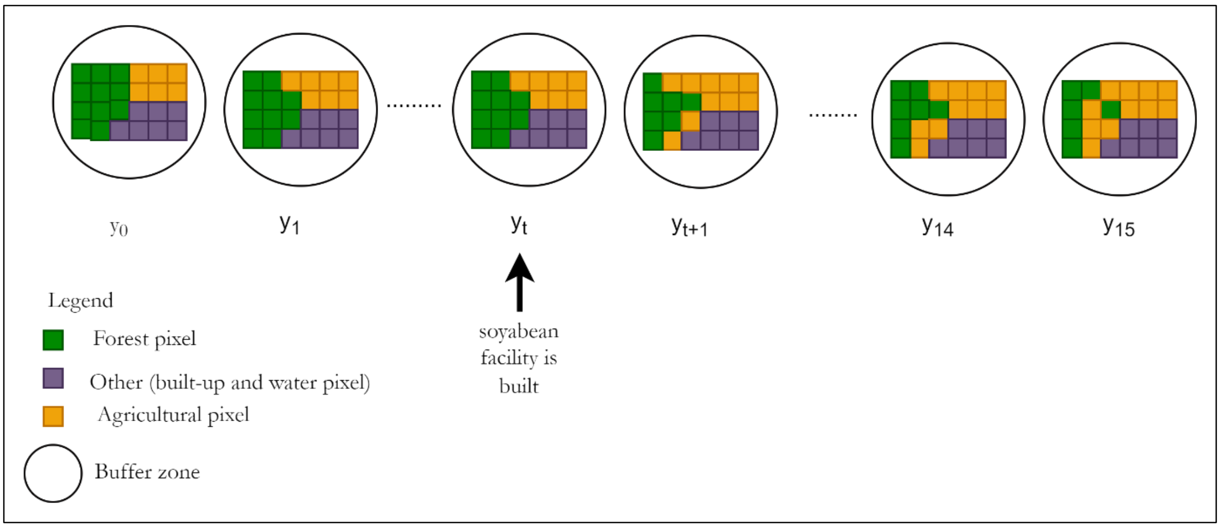


Figure 5: Pixel level demonstration of land cover change detection before and after soyabean grain facility.

The paired sample t-test formula is as below:

$$t = \frac{(\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}} \qquad \text{(Ross \& Willson, 2017)}$$

Where:

- t is the t-statistic.
- μ_1, μ_2 are the mean of the deforested area before and after the soyabean grain storage facility respectively.
- S_1^2, S_2^2 are the standard deviations of deforested area before and after the soyabean grain storage facility respectively.
- N_1, N_2 are the the number of buffers before and after the soyabean grain storage facility respectively.

In this context, if the calculated t-statistic was greater than the critical value at a significance level of 0.05, the null hypothesis was rejected, signifying a substantial change in land cover due to the soyabean grain storage facility.

2.4.2. Impact assessment of soyabean grain storage facility on deforestation

The figure below summeries the work flow for assessing the impact assessment of soyabean grain storage facility on deforestation in Mato Grosso.

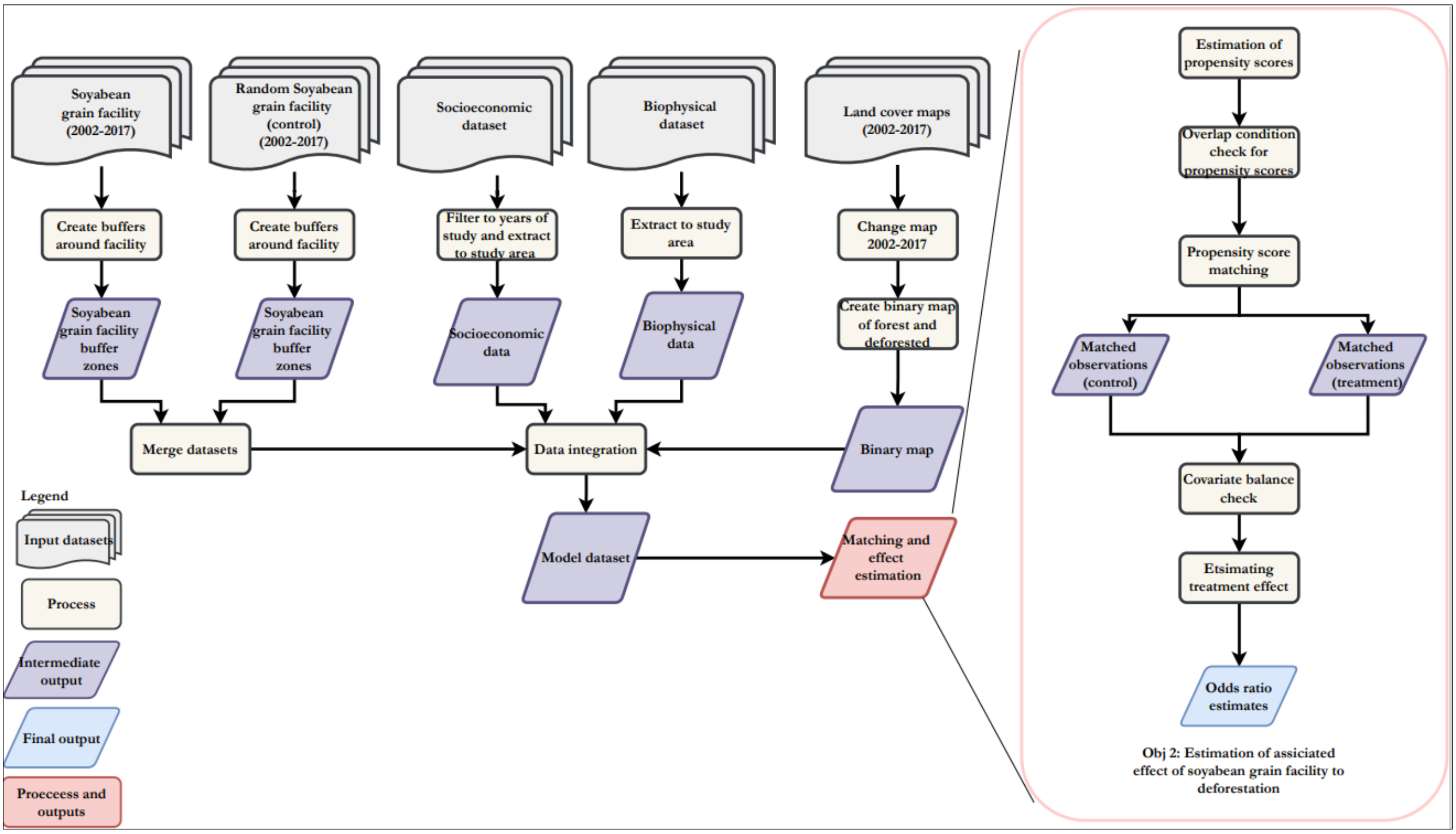


Figure 6: Methodology workflow for the second objective.

Treatment and control group creation and buffer definition

The location of soyabean grain facilities is not random; they are usually concentrated in areas with high agricultural activity. This clustering near active farming zones needs to be taken into account when evaluating their impact on deforestation, as it may influence the results (da Silva et al., 2021). As such, analysing their association to deforestation implies careful consideration of confounding factors namely socioeconomic and biophysical factors.

The matching technique helps mitigate the effects of non-random treatment allocation by mimicking the conditions of a randomized controlled experiement where the treatment and control groups are similar in terms of observed characteristics was used (Austin, 2011). In this study, this will be done by selecting control sites (buffers without soyabean grain facilities) that closely resemble treated sites (buffers with soyabean grain facilities) based on observable characteristics. We used the matching approach to account for various socio-economic, biophysical, and climatic factors. This involved aligning (pairing) treated and control groups to provide a robust evaluation of the effects of soyabean grain facilities on deforestation. Below is an illustration of control and treatment buffers:

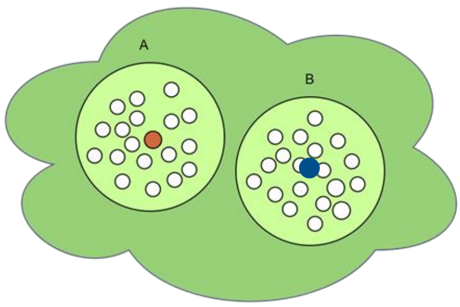


Figure 7: Demonstration of treatment (A) and control (B) buffers and random points.

To create the control group, random points within a 10 x 10 fishnet grid were generated. In each grid cell, 4 random points were generated mimicking soyabean storage facility location points. To this respect, a total of 245 points were generated to mimic buffer points locations. Buffers of 5, 10, and 25 km were drawn around these points in each group to define zones of influence. Within these generated buffers, points were randomly generated using the formula proposed by Krejcie & Morgan, (1970) given below:

$$\text{Sample size} = \frac{N.X^2.p.(1-p)}{e^2(N-1)+X^2.P(1-P)}$$

where:

- N is the size of the population (pixels within the buffer) from which the sample will be collected.
- $.X^2$ is the Chi-square for the confidence level (95% was used here) at 1 degree of freedom which is 3.841
- e is the margin of error measuring the desired level of accuracy (2.5% was used here).
- p is the proportion of the population which was set to 0.5 to ensure the estimated sample size is large enough to achieve the desired level of accuracy as suggested by Krejcie & Morgan, (1970).

This sampling equation ensures that the sample of forest pixels is representative of the entire population of forest pixels inside each buffer zone.

Following the generation of required sample size (points) for each buffer, pixel values of the socioeconomic and environmental variables were extracted to these points using ArcMap's "Extract to Points" feature. The "Spatial Join" feature was used to combine these datasets of values for each observable characteristic into a dataframe including only points with valid forested (1) or deforested (0), excluding points falling on the masked-out areas from the reclassified land cover change map. This land cover change map was generated by multiplying the 2002 binary mask with the 2017 land cover map. In the map all non-forest classes were categorized as deforested, and forest classes as forested.
Below is the sequence of images illustrating the methodology for generating control buffer points (A) using the grid, generating random points (B) within a buffer zone.

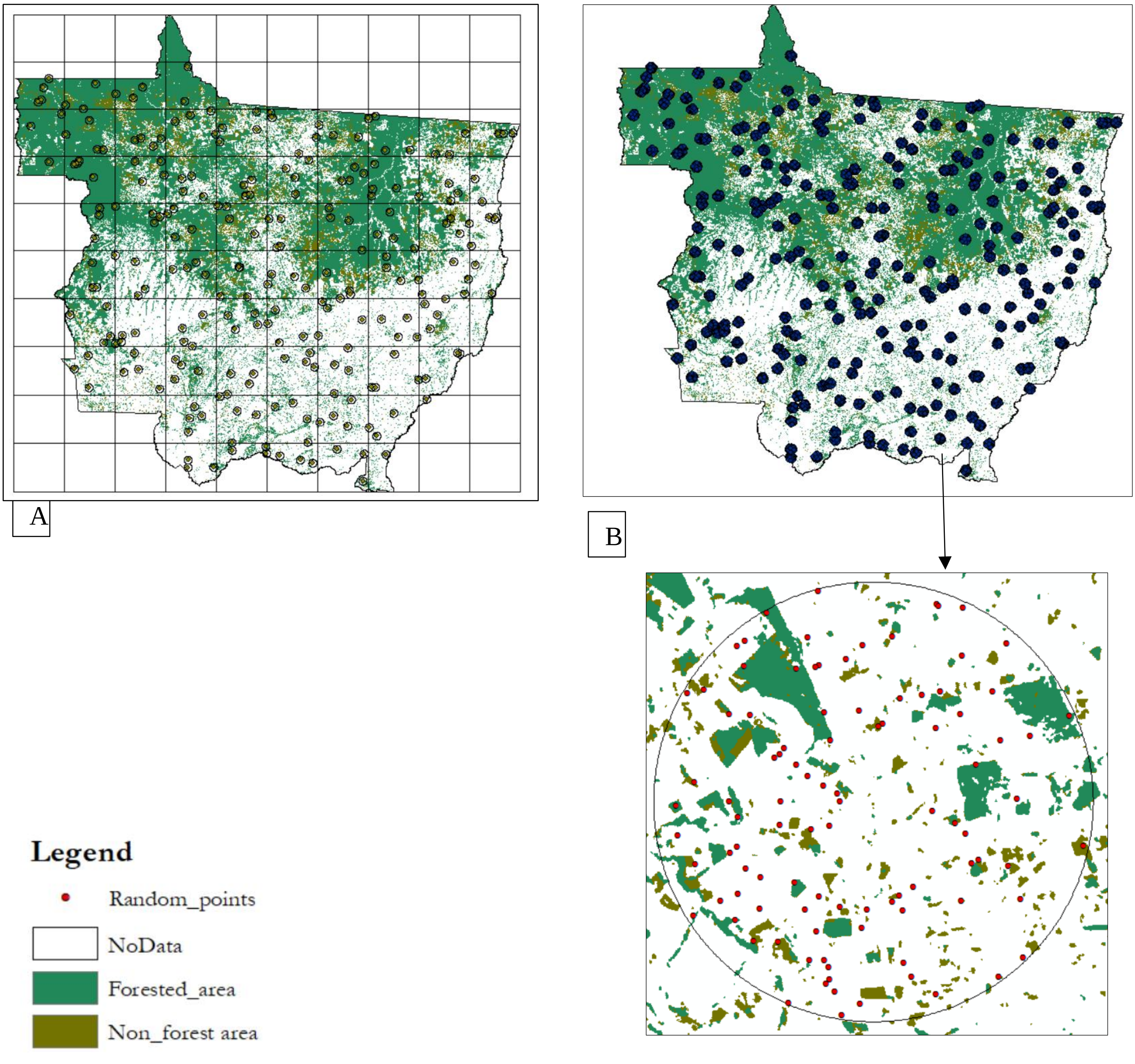


Figure 8: Demonstration of randomly generated points within the control buffer.

Note: The process of generating random points with buffers will be repeated for the treatment group as well for all buffer zones (5km, 10km and 25km).

Selection of observable variables for modelling

To ensure model robustness, we used the Variance Inflation Factor (VIF) method to detect and remove highly collinear covariates, thus refining our model (Shrestha, 2020).

Propensity score calculation

Following the integration of datasets and the joining of the two dataframes (control and treatment group), propensity scores using logistic regression were computed. A propensity score (π) for a feature (i) is defined by Rosenbaum and Rubin (1983) as the conditional probability (P) of assigning a participant to a particular treatment or comparison group (T) given a set of covariates(X). In this study, these scores represented the probability of an area having a soyabean grain storage facility given covariates i.e., socioeconomic, biophysical characteristics, and climatic conditions.

In this logistic regression model, the binary outcome variable is the presence or absence of soyabean grain facility presence, and the predictor variables include covariates related to soyabean grain storage facilities and potential confounders. This step ensures that each buffer has an assigned probability of being in the treatment group based on its covariate values. The logistic regression model estimates these scores for each buffer area and follows this formula:

$$\text{logit (soyabean facility presence)} = \ln\left(\frac{\text{soyabean facility presence}}{(1-\text{soyabean facility presence})}\right) = \alpha + \beta_1 * \text{gdp\_capita} + \beta_2 * \text{human\_(influence\_index)} + \beta_3 * \text{travel\_time} + \beta_4 * \text{elevation} + \beta_5 * \text{aspect} + \beta_6 * \text{slope} + \beta_7 * \text{soil\_(bulky\_density)} + \beta_8 * \text{soil\_depth} + \beta_9 * \text{soil\_pH} + \beta_{10} * \text{soi〚l\_organic〛\_(carbon\_stock)} + \beta_{11} * \text{precipitration} + + \varepsilon$$

Where:

D is the probability of d

α is the intercept term,

β_i are the coefficients for the covariates.

***Note**: Treatment is a binary variable indicating whether the buffer area contains a soyabean grain storage facility (1) or not (0), Covariates represent various factors such as socioeconomic indicators, biophysical characteristics, and climatic conditions.*

After calculating the propensity scores, overlap in these scores between treated buffers (those with soyabean grain storage facilities) and control buffers (those without such facilities) by assessing the overlap in propensity scores between the treatment and control group. This step is essential to confirm the common support, which means that there are comparable units in both groups. This allows for the creation of matched samples by pairing treated and control units with similar or close propensity scores during matching (Stuart, 2010).

Matching techniques

To enhance comparability and reliability in estimating the treatment effect of soyabean grain storage facilities on deforestation, nearest neighbour matching with replacement was employed. This technique was chosen for its ability to decrease bias by allowing controls that closely resemble multiple treated individuals to be used multiple times, particularly useful when few comparable control units are available. It simplifies the matching process by ensuring the order of matching does not affect the results and handles limited controls effectively. A bandwidth of 0.25 times the standard deviation of the logit of the propensity scores was used to reduce bias further, as recommended by Stuart (2010).

Post-matching analysis

After matching, post-matching analysis to reassess the balance of covariates between treated and control. This involved checking propensity score distribution post matching as well as standardized mean differences (SMD) to ensure that covariate distributions are similar across groups (Nguyen et al., 2017). Points that are not matched are dropped for the logistic regression of deforestation. This step ensures that the matched samples are comparable and that any observed differences in outcomes can be attributed to the treatment effect rather than to pre-existing differences between differences in groups.

Logistic regression and odds ratios

By first matching to ensure balance and then running a logistic regression on the matched sample, the study combined the strengths of both techniques. Matching reduces confounding by balancing covariates while regression adjusts for any remaining imbalance and estimates the effect of the treatment while controlling for other factors. This approach allowed for the computation of odds ratios, revealing the relationship between various predictors and the likelihood of deforestation. A logistic regression was applied to the matched sample to estimate the treatment effect of covariates on the outcome (deforestation). This method provided insights into the effects of the identified predictors while controlling for the matching factors. The odds ratios were computed to reveal the relationship between each variable and the likelihood of deforestation occurring. The equation below illustrates the logistic regression for the estimation of soyabean effect on deforestation.

$$\text{logit (D)} = \ln\left(\frac{\text{D}}{(1-\text{D})}\right) = \alpha + \beta_i * v_i$$

Where:

D is the probability of deforestation,

α is the intercept term,

β_i are the coefficients for the covariates.

Chapter Three: Study result

3.1. Transition matrix: Land cover dynamics in Mato Grosso between 2002 and 2017.

The table shows transitions between classes for the entire state of Mato Grosso.

Table 2: Mato Grosso land cover transition matrix (2002 to 2017).

		Land cover area (sq.km) in 2017																
	From Class	Cotton	Forest	Forest Plantation	Grassland	Mining	Mosaic of Crops	Mosaic of agriculture and pasture	Other non-vegetated areas	Pasture	River, Lake & Ocean	Rocky outcrop	Savanna	Soyabean	Sugar Cane	Urban Infrastructure	Wetland	Total
Land cover area (sq.km) 2002	**Cotton**	0.00	0.00	0.00	0.00	0.00	28.45	0.00	0.00	0.00	0.00	0.00	0.00	2755.31	0.00	0.00	0.00	2783.77
	Forest	0.00	0.00	67.62	999.40	30.86	2740.67	2338.76	181.81	77119.95	526.16	0.22	6866.26	650.58	10.80	7.53	999.53	92540.14
	Forest Plantation	0.00	18.35	0.00	2.13	0.04	24.94	10.47	0.51	27.81	0.08	0.00	19.04	10.75	1.17	0.01	0.29	115.59
	Grassland	0.00	750.76	40.83	0.00	27.26	665.24	1557.67	197.64	2563.56	1414.98	0.24	3126.54	1165.76	91.11	3.72	53274.44	64879.76
	Mining	0.00	4.17	0.10	2.83	0.00	0.34	0.26	0.05	6.47	37.24	0.00	1.19	0.06	0.00	0.18	1.93	54.83
	Mosaic of Crops	12.06	322.79	406.76	325.88	0.93	0.00	657.70	119.68	8910.40	9.74	0.01	365.60	63004.07	832.62	40.16	199.87	75208.26
	Mosaic of agriculture and pasture	0.16	627.48	135.69	525.72	2.19	2273.50	0.00	180.58	17106.11	326.34	0.22	4145.92	9261.45	44.09	13.37	477.71	35120.54
	Other non-vegetated areas	0.00	22.18	9.94	96.64	2.09	267.15	217.84	0.00	541.97	38.99	0.00	157.16	524.67	0.21	42.84	23.04	1944.73
	Pasture	0.10	22722.13	621.03	1636.97	67.89	36811.80	15698.86	330.00	0.00	159.37	0.06	14734.35	15999.44	427.23	137.14	1264.95	110611.33
	River, Lake & Ocean	0.00	530.19	0.19	1699.99	49.68	7.29	332.52	55.49	107.94	0.00	0.00	316.02	4.69	0.15	1.37	10347.85	13453.38
	Rocky outcrop	0.00	0.18	0.00	0.29	0.00	0.01	0.21	0.00	0.05	0.00	0.00	0.45	0.02	0.00	0.00	0.00	1.21
	Savanna	0.01	5746.38	105.75	2389.97	21.50	865.45	8544.82	680.19	23774.79	350.77	0.42	0.00	2379.38	18.05	43.85	2375.73	47297.08
	Soyabean	3062.24	128.04	53.92	57.24	0.27	39603.66	227.52	36.38	1441.98	7.18	0.03	161.35	0.00	212.39	22.23	14.37	45028.79
	Sugar Cane	0.00	3.21	0.46	26.88	0.00	402.30	11.58	0.56	187.30	0.17	0.00	6.98	105.99	0.00	0.00	82.23	827.66
	Urban Infrastructure	0.00	0.39	0.01	0.22	0.02	3.40	1.36	1.71	4.89	1.01	0.00	2.29	2.21	0.00	0.00	0.39	17.90
	Wetland	0.00	816.37	0.85	54076.71	4.63	540.16	867.16	31.12	1846.73	8850.80	0.00	2774.40	69.33	144.62	5.66	0.00	70028.54
	Total	3074.57	31692.62	1443.16	61840.87	207.36	84234.37	30466.74	1815.73		11722.84	1.21	32677.53	95933.72	1782.45	318.06	69062.32	559913.49

From the table above, forest is seen to convert to agricultural related classes namely mosaic of agriculture and pasture, pasture and soyabean. Land cover change (deforestation) for the period 2002 to 2017 can be highly associated to agricultural activities over the period.

Besides the presence of zeros in certain cells of the transition table above which indicates that there was no transition between those specific land cover types from 2002 to 2017, there exists unusual transition i.e., forests appear to be converting to river, lake, and ocean areas. This could be due to pixel-based classification and changes in river courses. The classifier likely assigns pixel classes based on the majority class within each pixel. Comparing Google Earth Pro images with classified maps from 2003 and 2012 confirms these observations.

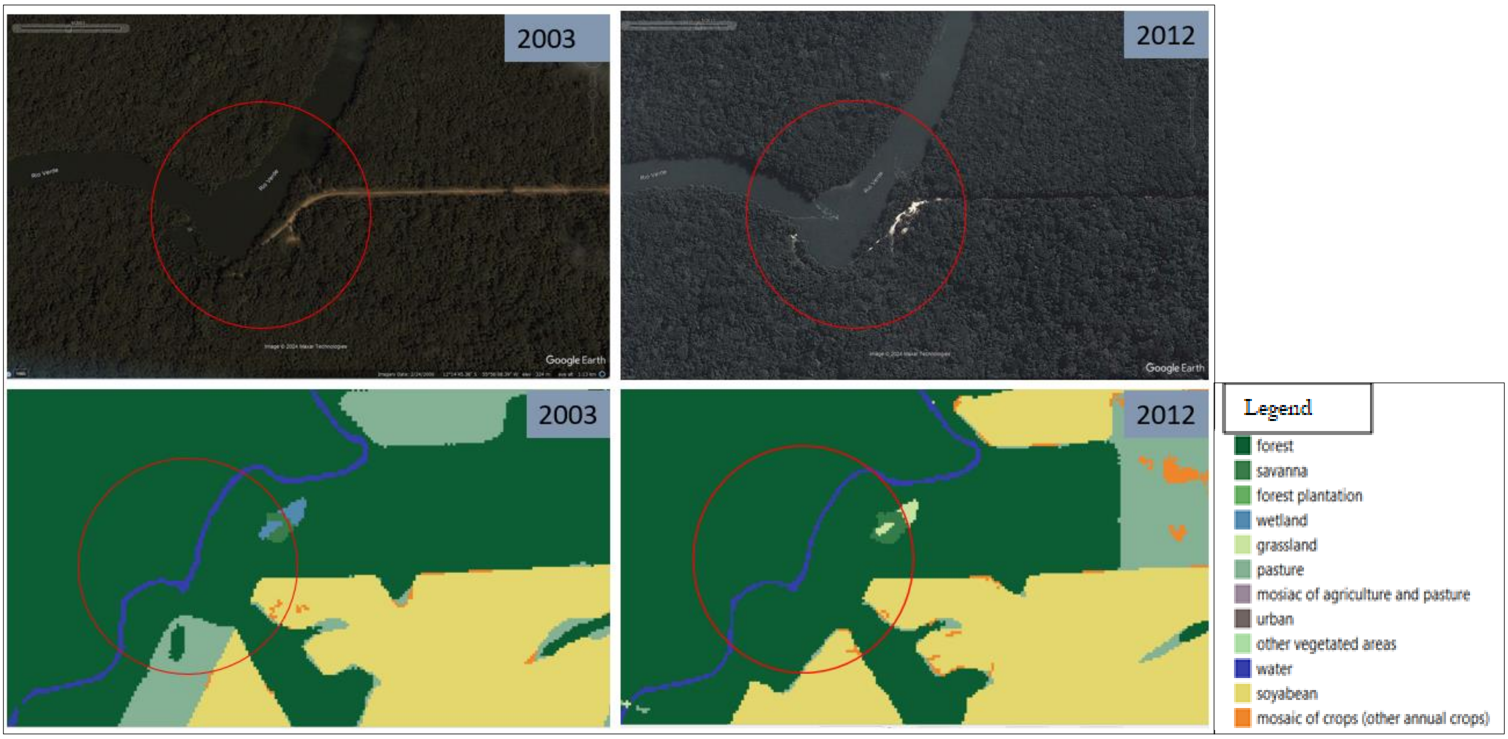


Figure 9: Observed misclassifications.

3.2. Deforestation hotspot and spatial extent

The figure below shows a cumulative deforestation map showing areas where deforestation has occurred consistently throughout the study period. The soyabean grain storage facility data was overlaid on the deforestation mask to check for spatial overlaps between areas of forest loss and the locations of these facilities. It was observed that there is a spatial overlap between the facilities and the deforested areas which can be indicative of the influence of the facilities as well as increased agricultural activities on deforestation.

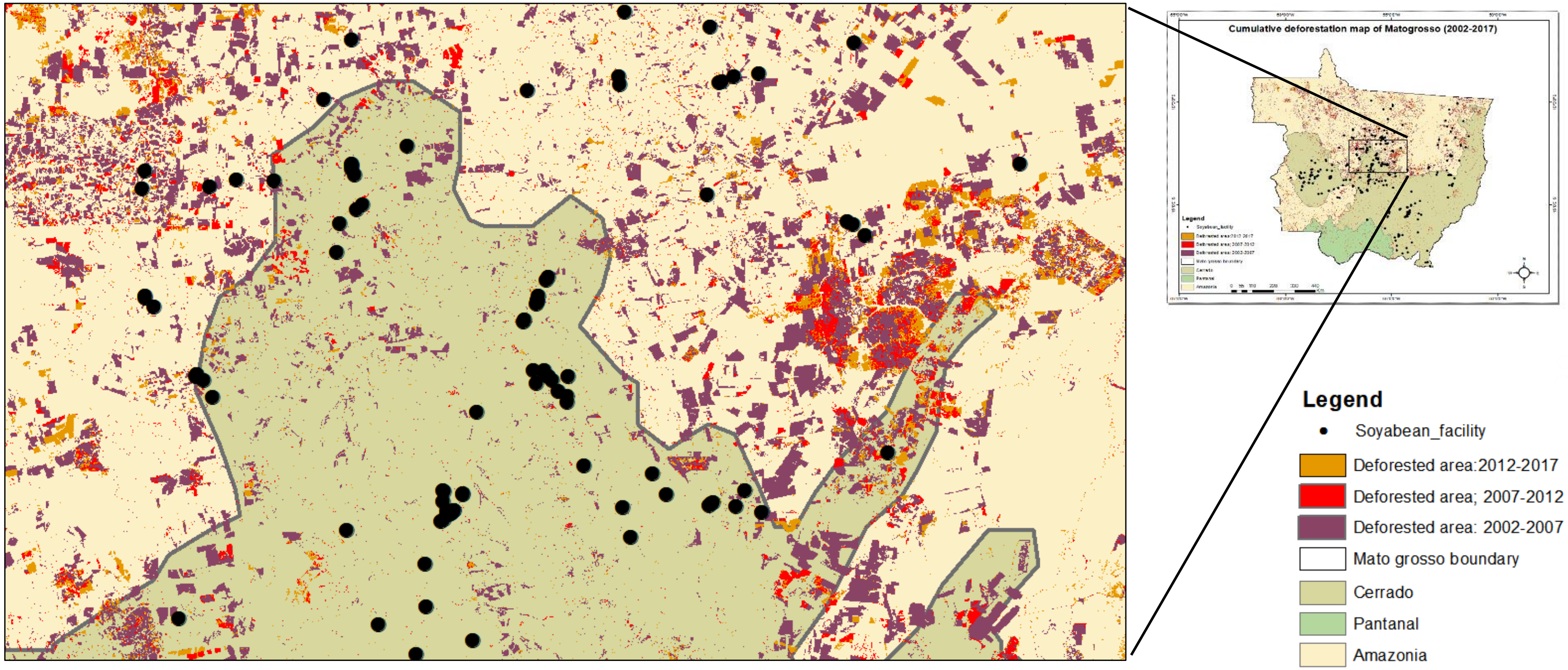


Figure 10: Demonstration of deforested areas from 2002 to 2017.

Note that between 2002 and 2007, 159 soyabean facilities were built, between 2007 and 2012, 137 were built and 73 were built between 2012 and 2017.

3.3 Association of land cover changes with soyabean facilities in Mato Grosso

3.3.1. Mean annual deforestation rates within buffers in Mato Grosso

A temporal trend analysis of mean annual deforestation rates within the buffer zones from 2002 to 2017 was done to assess the evolution of deforestation in the context of soyabean facility establishment.
The figure below consists of three-line graphs, each depicting the mean deforestation rate (%) between 2002 and 2017 for the buffer zones are 5 km, 10 km, and 25 km. Generally, all three graphs show a significant decline in deforestation rates from 2003 to around 2006-2008 after which the rates stabilize at a much lower level. The 5 km buffer zone shows the highest initial deforestation rate of around 8% in 2003, which drops steeply to around 1% by 2006. The rate fluctuates slightly but remains relatively low in subsequent years. The 10 km buffer zone starts with a deforestation rate of around 6% in 2003, which also drops significantly to about 1% by 2006. The rate then shows minor fluctuations around 1%. While the 25 km buffer zone begins with a deforestation rate of approximately 5% in 2003 and drops to around 1% by 2006. This buffer zone also shows minor fluctuations around 1% in the following years.

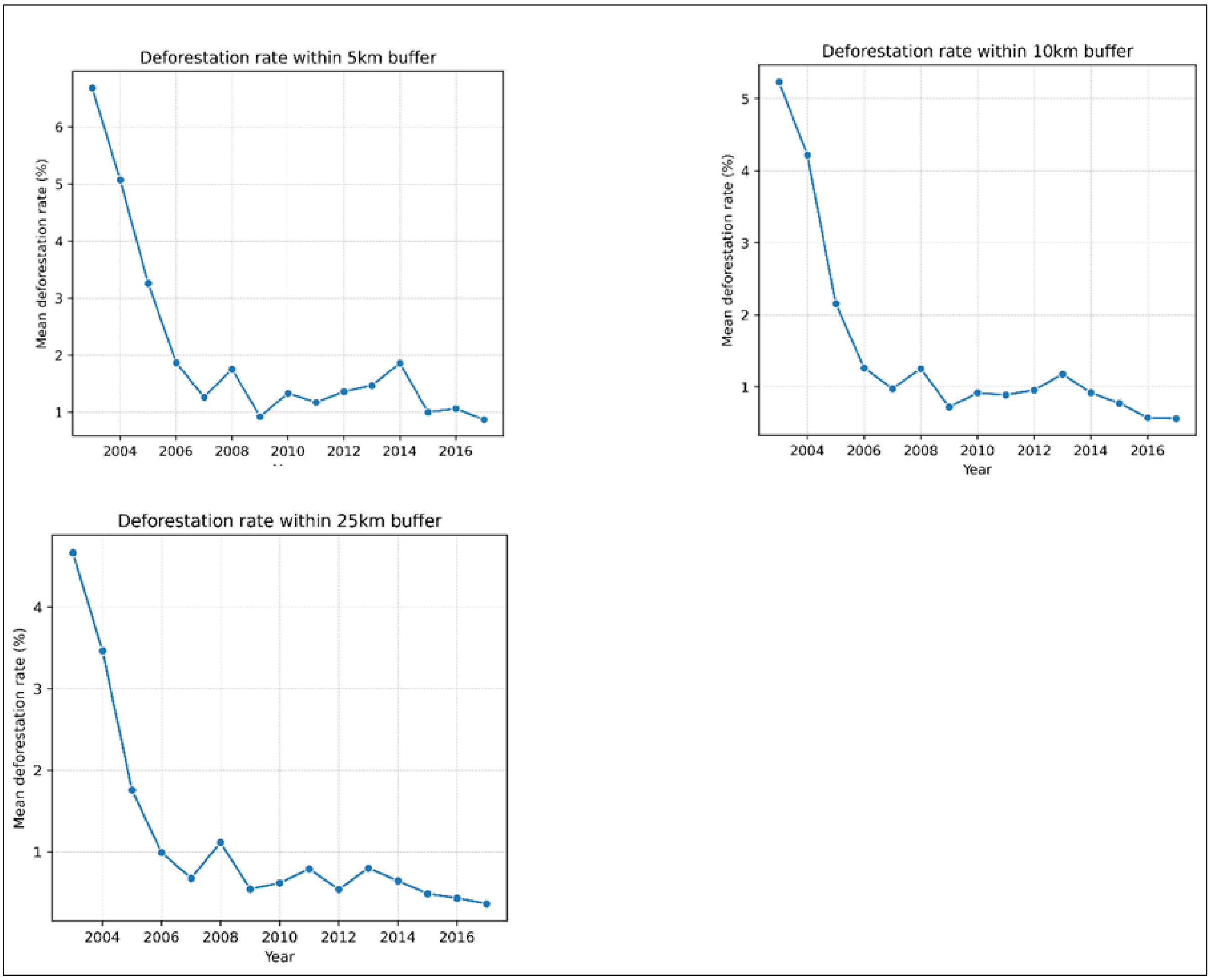


Figure 11: Trendlines of mean deforestation rates within buffers (5km, 10km and 25km) from 2002 to 2017.

3.3.2. Spatial impact assessment of soyabean facilities on deforestation rates and land cover change.

Building on the temporal assessment, we restricted the timeframe to compare mean deforestation rates for three years before and after the establishment of each soyabean grain facility. This focused analysis reveals how land cover change varies spatially around soyabean facilities over time, indicating whether these facilities significantly affect deforestation rates.

The figure below shows box plots that compare average deforestation rates three years before and after the construction of soyabean grain facilities within 5 km, 10 km, and 25 km buffer zones. The box plots show that the mean deforestation rates decreased from 1.57% to 1.16% in the 5 km buffer, 1.09% to 0.76% in the 10 km buffer, and 0.97% to 0.58% in the 25 km buffer. The statistical analysis shows that the decrease in deforestation rates is significant for all the buffer zones, with t-statistics of 4.09, 5.64 and 8.54, respectively, and p-values of 0.00. This suggests that construction of soyabean grain facility is associated with a significant reduction in deforestation rates in the defined buffer zones.

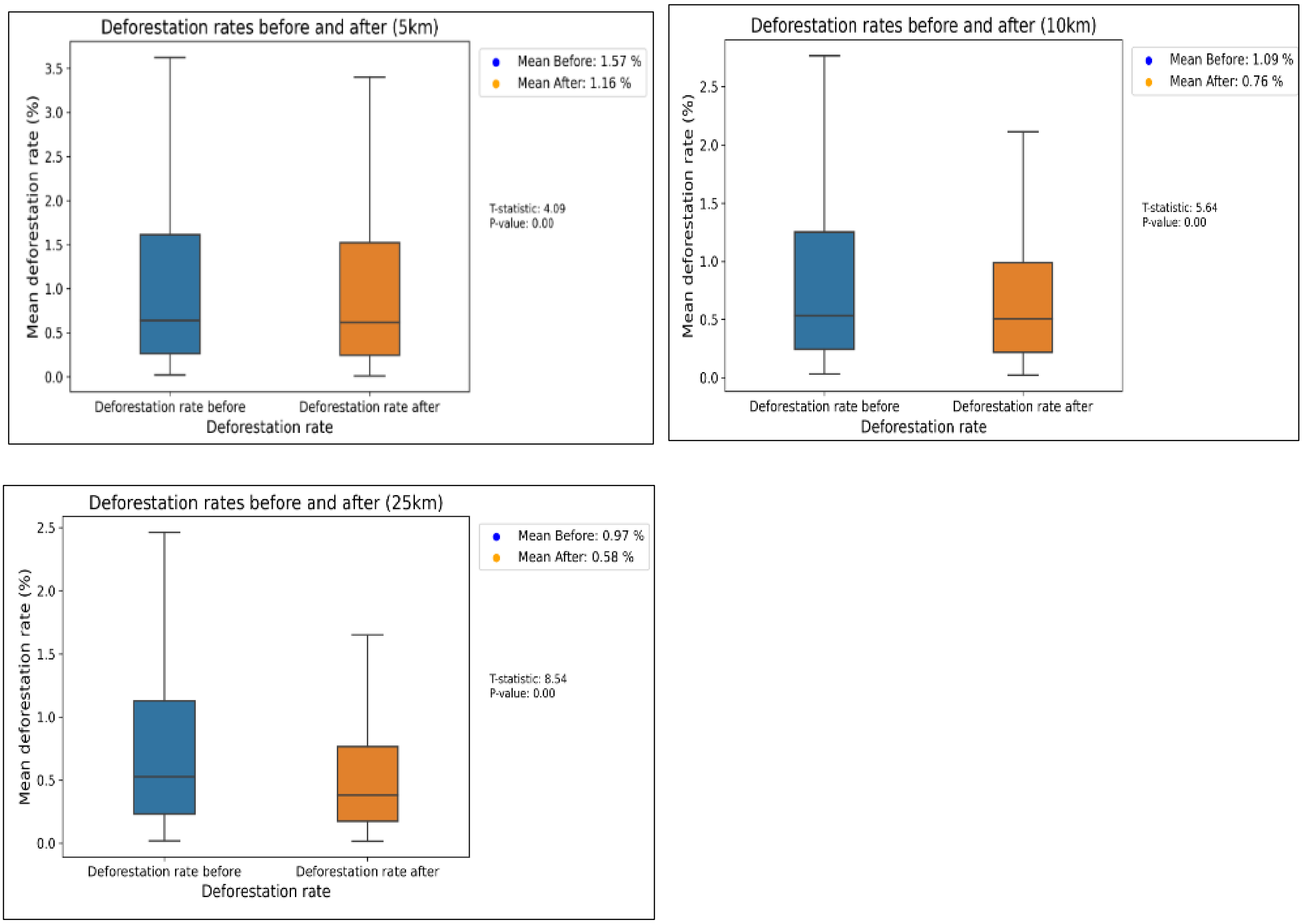


Figure 12: Boxplots of deforestation rates before and after the establishment of soyabean grain storage facilities at different buffer zones (5 km, 10 km, and 25 km).

3.3.3. Spatial land cover changes around soyabean storage facilities (3 years pre/post establishment).

To further understand the association of soyabean grain facilities on land cover change, we assessed whether the establishment of these facilities significantly affects the area and nature of land cover transition. The following figures present the evaluation results focusing on specific land cover types that forest areas transitioned into namely pasture, soyabean, mosaic of crops, and mosaic of agriculture and pasture. This analysis provides a detailed perspective on how different land cover types are influenced by the proximity to soyabean grain facilities, highlighting the environmental impacts of these developments.

5 km buffer zones

The figure examines land cover changes within a 5km buffer zone around soyabean grain storage facilities from 2002 to 2017. The analysis reveals significant reductions in pasture area (mean area before 0.64 km², mean area after: 0.29 km², T-statistic: 4.09, P-value: 0.00) and mosaic of agriculture and pasture area (mean area before: 0.02 km², mean area after: 0.02 km², T-statistic:1.25, P-value: 0.21), indicating a shift in land use patterns. Similarly, the mosaic of crops area shows a decrease (mean area before 0.06 km², mean area after: 0.05 km², T-statistic: 1.33, P-value: 0.19), suggesting that these zones are experiencing significant land cover transformations. However, the soyabean area exhibits a slight, non-significant increase (mean area before 0.02 km², mean area after: 0.02 km², T-statistic: 0.32, P-value: 0.75), indicating that soyabean cultivation might not have expanded as rapidly in the immediate vicinity of storage facilities.

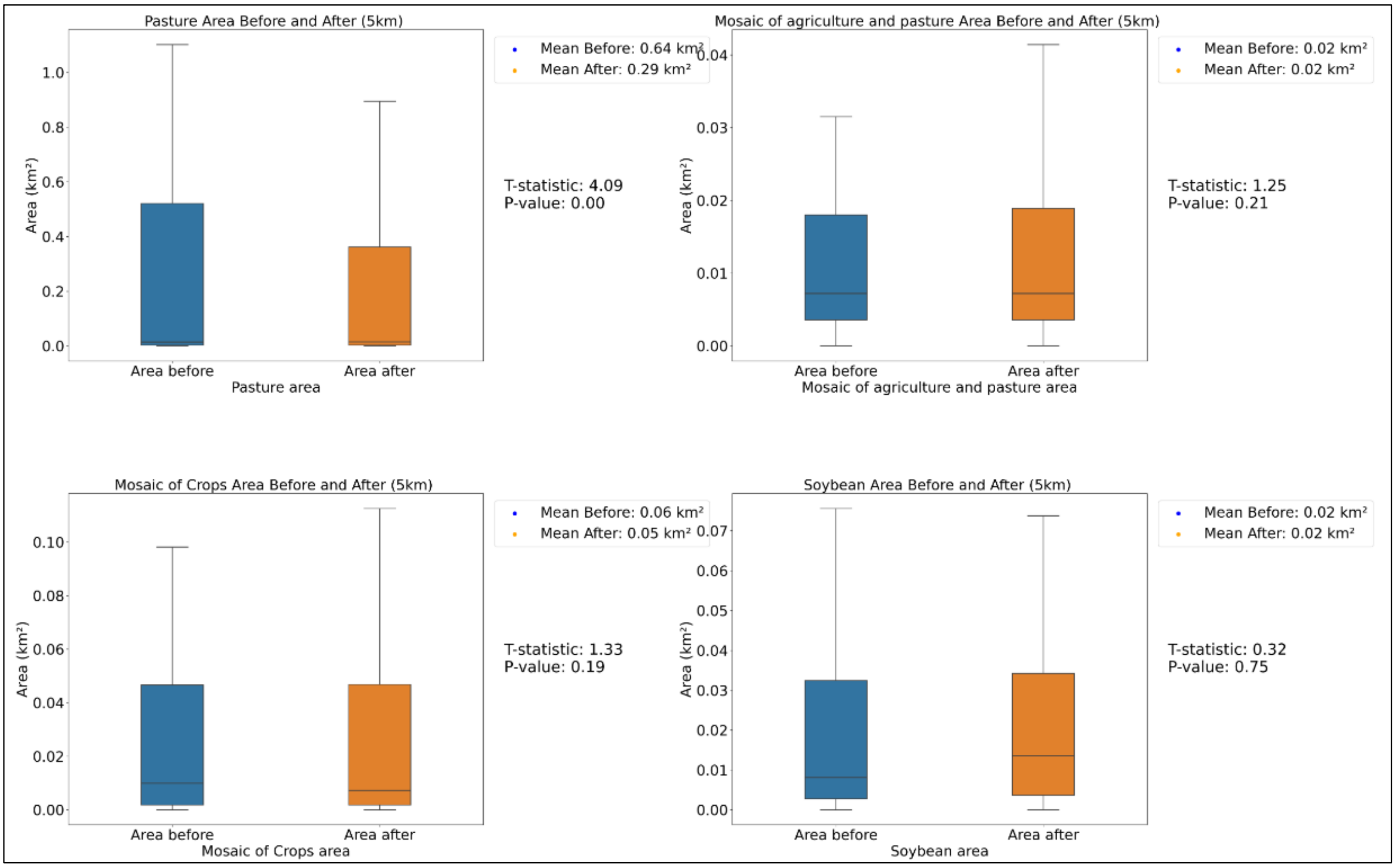


Figure 13: Box plots of mean area distributions (in km²) for specific land cover types before and after the establishment of soyabean grain storage facilities at different buffer zones within 5 km.

a. 10 km buffer zones

The figures below show similar analysis extended to the 10km buffer zone. It reveals more pronounced land cover changes. The pasture area shows a significant reduction (mean area before: 2.45 km², mean area after: 1.21 km², T-statistic: 4.57, P-value: 0.00), which is more substantial compared to the 5km buffer. The mosaic of agriculture and pasture area also decreases significantly (mean area before 0.11 km², mean area after: 0.08km², T-statistic: 1.96, P-value: 0.05). The mosaic of crops area reduction is still significant (mean area before: 0.24 km², mean area after: 0.18 km², T-statistic: 2.27, P-value: 0.02). Interestingly, the soyabean area change remains non-significant (mean area before 0.11 km², mean area after: 0.09 km², T-statistic: 2.08, P-value: 0.04). These patterns suggest that land cover changes are more pronounced as the buffer zone increases to 10km.

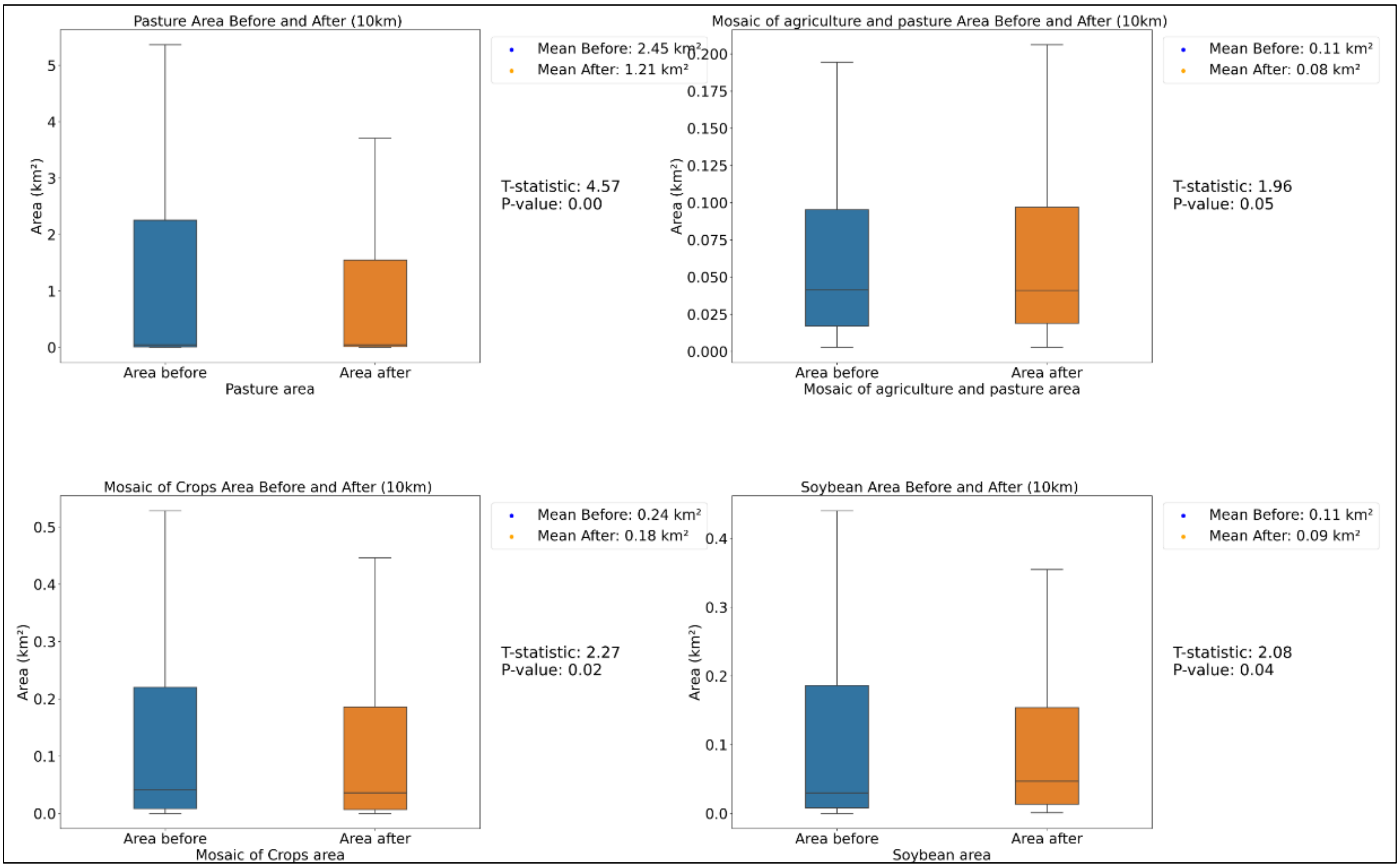


Figure 14: Box plots of mean area distributions (in km²) for specific land cover types before and after the establishment of soyabean grain storage facilities at different buffer zones within 10 km.

.

b. 25km buffer zones

The figures below show similar analysis further extended to the 25km buffer zone. The reduction in pasture area is notably large (mean area before: 19.57 km², mean area after: 9.76 km², T-statistic: 5.70, P-value: 0.00), reflecting extensive land use changes over a larger area. The mosaic of agriculture and pasture area also shows a significant decrease (mean area before: 1.35 km², mean area after: 0.78 km², T-statistic: 5.20, P-value: 0.00), as does the mosaic of crops area (mean area before: 1.84 km², mean area after:1.48 km², T-statistic: 2.47, P-value: 0.01). The soyabean area shows a slight but significant reduction (mean area before 0.72 km², mean area after: 0.55 km², T-statistic: 3.24, P-value: 0.00). These findings indicate that the establishment of soyabean storage facilities has widespread effects on land cover changes extending beyond immediate proximities.

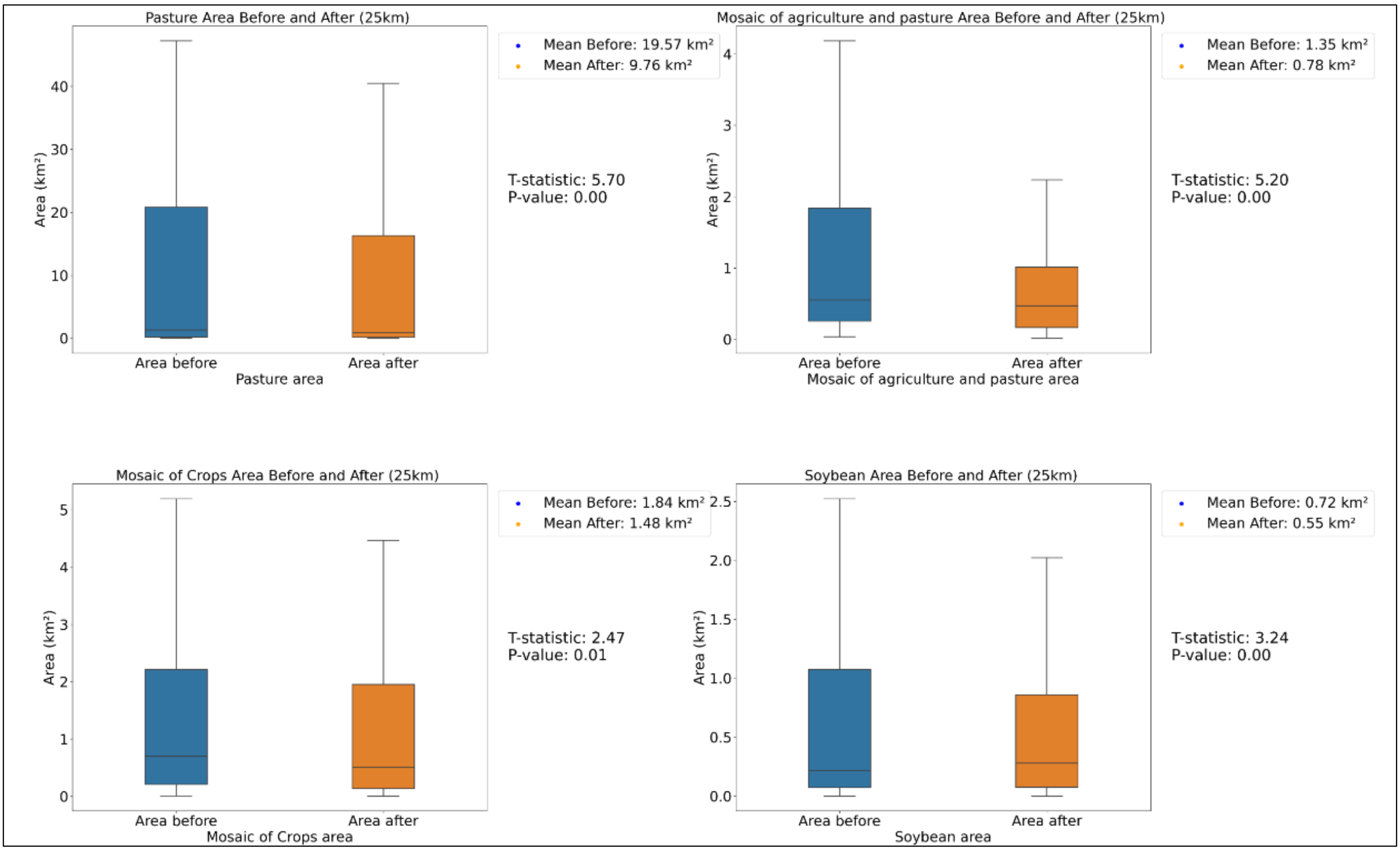


Figure 15: Box plots of mean area distributions (in km²) for specific land cover types before and after the establishment of soyabean grain storage facilities at different buffer zones within 25 km.

3.4. Impact assessment of soyabean grain facilities on deforestation

3.4.1. Assessment for multicollinearity between variables

Table 3 below shows the computed variance inflation factors (VIF) of the explanatory variables considered to influence deforestation. VIF value greater than 10 were excluded from the analysis. The VIF results indicate that multicollinearity is a significant issue for several variables across all the 3 buffer zones, namely average precipitation, bulky density, percent clay, soil pH, Soil depth, soil organic carbon and elevation show collinearity and thus were excluded from further analysis. Below are the computed VIF.

Table 2: Variance inflation factors.

Variables	**Variance inflation factors (VIF)**		
	5km buffer	**10km buffer**	**25km buffer**
Aspect	3.59	3.581	3.55
Elevation	19.81	19.40	20.74
Slope	2.10	2.142	2.132
Soil_depth	3088.83	2522.12	2345.78
Soil_organic_carbon _content	11.88	12.39	13.35
Percent_clay	30.38	30.09	30.19
Bulk_density	1296.44	1288.44	1233.76
Soil_pH	558.79	559.60	593.52
GDP_Capita	5.61	5.19	5.09
Travel_time	4.05	4.041	4.09
Human_Influence index	2.91	2.974	2.94
Average_precipitation	198.10	198.38	215.82
Forest_km ^{2}	1098.43	555.39	277.84
Soyabean_facility	3.03	3.05	3.12

3.4.2. Propensity scores estimation

The propensity scores for the treated and control groups show adequate overlap across all buffer zones: 5 km (0.0235 to 0.9919), 10 km (0.00675 to 0.99967), and 25 km (0.0045 to 0.99961). This overlap ensures that the treatment and control groups are comparable, which is essential for making valid causal conclusions (see appendix 3).

3.4.3. Performing matching balanced

Figure 15 below shows the density distributions of propensity scores for the treated group (buffers with soyabean grain facility) and the control group (buffers without soyabean grain facility) before and after nearest neighbour matching with replacement show successful balancing across all buffer sizes (5 km, 10 km, 25 km). Initially, the treated group has higher propensity scores while the control group has lower scores with some crucial overlap in the middle range. After matching, the distributions of both groups are very similar. This indicates that the matching process has effectively created comparable groups which enhances the validity of the analysis of the treatment effect by reducing bias. In addition, we computed Standardized Mean Differences (SMDs) to further check the balance between covariates, with a threshold of less than 0.1 indicating acceptable balance (see appendix 4).

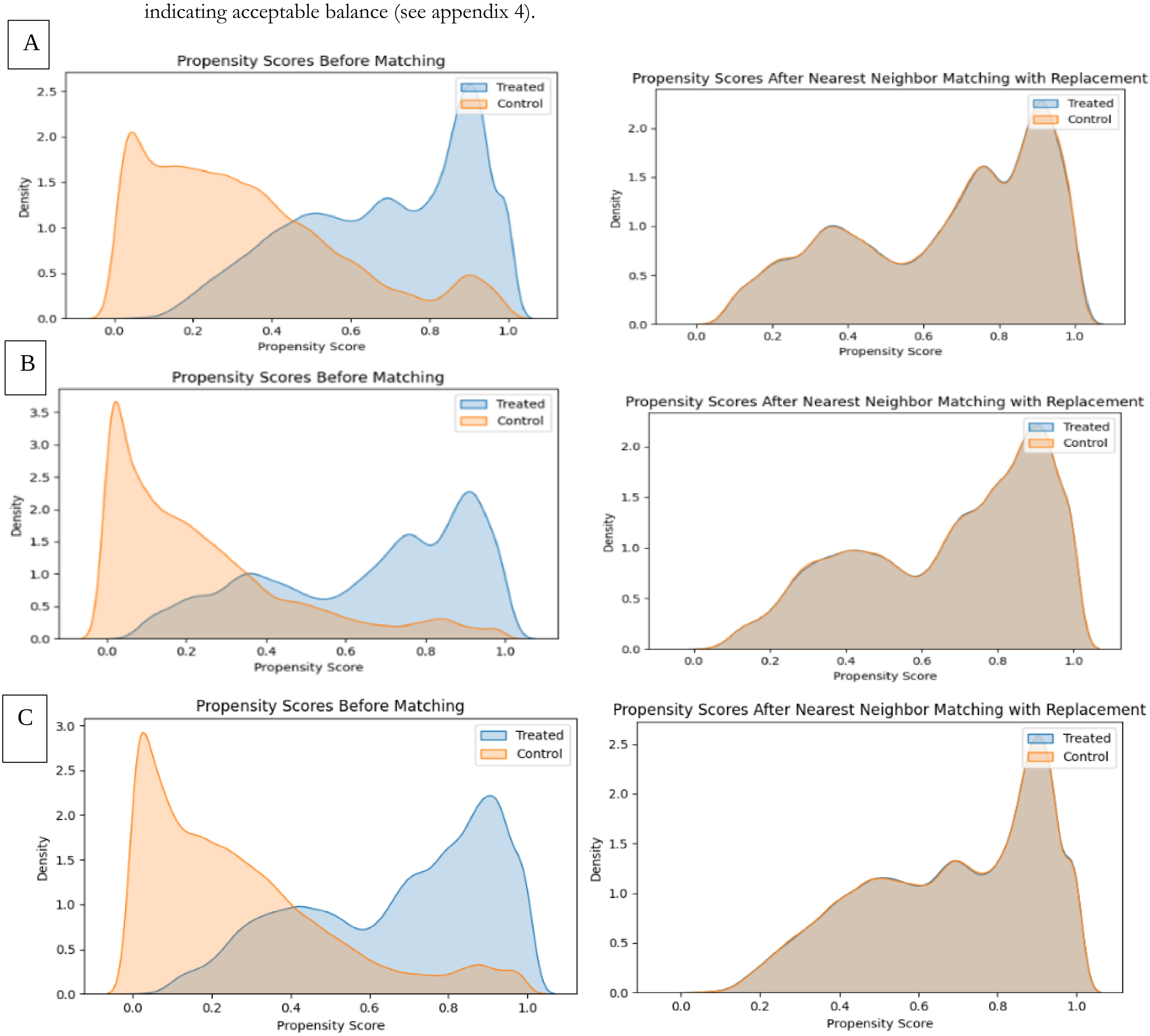


Figure 16: Pre and post matching density distributions plots of propensity scores for the treated group) and the control group before and after nearest neighbour matching with replacement across buffer sizes of 5 km (Panel A), 10 km (Panel B), and 25 km (Panel C).

3.4.4. Estimates of odds ratios on deforestation

Below are the results showing the odds ratios, which estimate the relative effect of the treatment (i.e., presence of soyabean grain storage facilities) compared to the control (i.e., absence of such facilities) while considering other variables in the model.

Across all buffer zones from 5 km to 25 km, the presence of soyabean facilities significantly increases the odds of deforestation with the impact decreasing as the buffer zone expands (odds ratios: 5 km - 2.20, 10 km - 1.77, 25 km - 1.47; p=0.000). Additionally, from 5km to 25km, aspect has a minimal positive impact on the likelihood of deforestation with odds ratios remaining close to 1 (p=0.000). Slope on the other hands consistently shows a substantial decrease in the likelihood of deforestation with odds ratios around 0.85 (p=0.000) across all buffers. GDP per capita consistently leads to a slight less reduction in deforestation with odds ratios between 0.98 and 0.99 (p=0.000) from 5km to 25km buffer zone. The study also shows that as travel time to the nearest city increases, there is a very slight and practically negligible increase in deforestation odds, with odds ratios around 1.0004 to 1.0006 (p=0.000) across all buffer zones. Lastly, the human influence index consistently indicates a slight decrease in deforestation odds from 5km to 25km buffer with odd ratios ranging between 0.97 to 0.99 (p=0.000).
The table below summarises the results.

Table 3: Odd ratio estimates.

	5km		**10km**		**25km**	
Variable	**Odds Ratio**	**p>|z|**	**Odds Ratio**	**p>|z|**	**Odds Ratio**	**p>|z|**
Const.	0.352654	0.000	0.371740	0.000	0.430606	0.000
Soyabean grain facility	2.199309	0.000	1.768814	0.000	1.471436	0.000
Aspect	1.000176	0.000	1.000305	0.000	1.000416	0.000
Slope	0.849650	0.000	0.846775	0.000	0.844263	0.000
GDP_Capita	0.984699	0.000	0.984809	0.000	0.990965	0.000
Travel_time	1.000628	0.000	1.000466	0.000	0.999946	0.000
Human influence index	0.986764	0.000	0.988400	0.000	0.969168	0.000

Chapter Four: Discussion of results

4.1. Temporal trends and spatial extent of deforestation (2002-2017)

A temporal trend analysis of mean annual deforestation rates from 2002 to 2017 within the 5 km, 10 km, and 25 km buffer zones around soyabean grain facilities reveals significant declines over time. Initially, the 5 km buffer zone had the highest deforestation rate at approximately 8%, followed by the 10 km zone at 6%, and the 25 km zone at 5% in 2003. By 2006, deforestation rates in all zones had dropped to around 1% and remained low thereafter. These findings align closely with broader state and national and regional temporal trends observed by various studies. Studies at national level by Soares-Filho & Rajão, (2018) and Macedo et al., (2012) focusing on the Brazilian Amazon observe similar trend and attribute the reduction in deforestation to a combination of policies and actions namely PPCDAm (Portuguese acronym for Action Plan for Deforestation Prevention and Control in the Legal Amazon) the soy moratorium, the blacklist, and the CAR (Cadastro Ambiental Rural, or Rural Environmental Registry) registry. A study by Graesser & Ramankutty, (2017) focusing on the slowdown of deforestation rates in Latin American besides Brazil observed similar temporal trend and attributed changes to shifts in agricultural practices and conservation efforts. The authors note that while soyabean expansion initially led to high deforestation rates, improved land use policies have helped reduce and stabilize these rates over time.

The results suggest that there is a higher initial decline in deforestation rates within the 5 km buffer compared to the 10 km and 25 km buffers. Despite focusing on these localized buffer zones, the trends observed in deforestation rates are consistent with broader patterns seen at national and regional levels. The areas closest to the soyabean facilities saw the fastest decline in deforestation rates initially, and these local trends mirror the general trends observed in larger-scale studies. This underscores the fact that areas closest to soyabean facilities are most affected by deforestation policies, highlighting the significant impact of these policies in mitigating deforestation near agricultural infrastructure. Furthermore, policy effects tend to be more pronounced in smaller buffers where deforestation rate reductions are more immediate and significant compared to larger buffers where the impact is more gradual.

4.2. Impact assessment of soyabean facilities on deforestation (2002-2017).

The paired t-test results suggest that the establishment of soyabean grain storage facilities is associated with a significant reduction in deforestation rates within the defined buffer zones, though this does not imply causality. The largest reduction (percentage point difference between the deforestation rates before and after establishment) is observed in the 5 km buffer then 10 km buffer and lastly 25km.

These findings align with studies by Hargrave & Kis-Katos (2013) which found that increased monitoring and enforcement efforts near areas of economic activity, such as soyabean plantations have contributed to reduced deforestation rates. Additionally, Soares-Filho & Rajão (2018) highlight the effectiveness of policies like the soy moratorium and CAR Registry in promoting sustainable practices and discouraging illegal deforestation over large areas. Furthermore, Morton et al. (2006) demonstrated that soyabean expansion initially led to high deforestation rates near new plantations but decreased as these areas became more established and monitored.

The spatial implications of these results suggest that the proximity to soyabean facilities significantly impacts deforestation rates. The sharpest decline within the 5 km buffer indicates that areas closest to soyabean storage facilities benefit the most from direct interventions such as improved monitoring, enforcement and sustainable practices implemented by agricultural producers. This immediate oversight becomes less effective as the distance increases evidenced by the smaller yet significant reductions in the 10 km buffer. In the 25 km buffer, the reduction in deforestation rates, while still significant, likely reflects broader policy enforcement and sustainable practices rather than direct influence from the facilities. These pattern highlights the varying degrees of access and effectiveness of interventions at different distances with the closest areas experiencing the most substantial benefits from direct oversight and interventions. These spatial dynamics underscore the importance of tailored policy efforts and monitoring strategies to ensure continued reductions in deforestation rates across different buffer zones.

4.3. Land cover changes around soyabean facilities (3 Years pre/post-establishment).

The study found that much of the deforestation was attributed to forest changing to agriculture land cover classes. The analysis of land cover changes within the 5 km buffer zone around soyabean grain storage facilities highlights significant transformations particularly the reduction in pasture areas. This suggests that land previously used for grazing is being repurposed due to the influence of the storage facilities. The stability observed in the mosaic of agriculture and pasture area indicates a balanced use of land for mixed purposes while the slight, non-significant increase in soyabean cultivation suggests limited expansion close to the facilities. These findings show the direct influence of storage facilities within the 5 km buffer effectively encourages sustainable practices and reduces the need for grazing land. Extending the analysis to the 10 km buffer zone, more pronounced land cover changes are evident. Significant reductions in pasture areas and mixed land use categories indicate that the influence of soyabean storage facilities extends beyond the immediate vicinity. This broader impact reflects the adoption of sustainable practices and improved land management over a wider area. The relatively stable soyabean cultivation in this zone suggests a balance between expanding agriculture and maintaining forested areas. The findings within the 10 km buffer demonstrate that the storage facilities' influence promotes sustainable land use practices across a broader landscape. Lastly, the 25 km buffer zone shows extensive land use changes, with significant reductions in pasture and mixed land use areas, indicating a substantial environmental influence of soyabean storage facilities over larger distances. This widespread adoption of sustainable practices is likely driven by effective policy enforcement and the presence of protected areas with stricter environmental regulations. The slight reduction in soyabean cultivation in this zone suggests a balanced approach to land use, preventing excessive deforestation while accommodating agricultural needs. The analysis across all buffer zones reveals that the establishment of soyabean storage facilities promotes sustainable land use practices and reduces deforestation, with the most substantial impacts observed within the 25 km buffer. This broad impact underscores the importance of integrated environmental management strategies for achieving long-term sustainability in agricultural regions.

These results on land cover changes across buffers align with the findings by Vieira et al., (2022) observed increased crop intensification practices and reduced cultivatable land within the same study period of 2002 to 2017. The study attributed the results to the broader enforcement of sustainable practices and market-driven shifts from grazing to crop cultivation. The stability in the mosaic of agriculture and pasture areas and the minor changes in soyabean cultivation within the buffer zones reflect a balanced approach to land management influenced by both policy enforcement and market conditions such as price.

4.4. Impact of soyabean grain facility on deforestation: Insights from socioeconomic, biophysical, and climatic factors

The presence of soyabean facilities significantly increases the odds of deforestation across all buffer zones. The impact is most substantial within the 5 km buffer and decreases with distance in the 10 km and 25 km buffers indicating that the influence of soyabean facilities on deforestation diminishes as distance increases. This trend is consistent with findings in studies by Morton et al., (2006) and Müller et al., (2011) supports the observation that industrial agricultural activities, particularly large-scale soyabean production have significant localized economic benefits but also drive extensive deforestation both locally and in neighbouring areas.

However, when looking at other factors like terrain slope, economic status, and human activity, we see more details that help explain these findings more precisely. For instance, the observed minimal impact of aspect suggests that the orientation of land does not significantly influence deforestation likelihood aligning with studies by Måren et al. (2015), which found that the orientation of land (aspect) has a minimal impact on deforestation likelihood (Måren et al., 2015). Travel time, a proxy to accessibility, shows a very slight decrease in deforestation odds further away from the soyabean facility. This suggests that increasing time to accessibility and logistical aspects do not substantially influence deforestation risk. Studies by Picoli et al. (2020) and Yang et al. (2021) highlight that easier accessibility to agricultural markets and other logistical infrastructure is one key driver for agricultural expansion. Additionally, increased GDP per capita correlates with reduced deforestation supporting the environmental Kuznets curve hypothesis where economic growth initially leads to deforestation but eventually promotes conservation (Crespo Cuaresma et al., 2017; Rosa et al., 2013). The human influence index shows a more significant reduction in deforestation with increasing distance from soyabean facilities aligning with findings by Lambin et al. (2001) that emphasize the role of managed and monitored areas in reducing deforestation.

These findings suggest that while soyabean facilities drive deforestation especially in proximity, other factors such as terrain slope, economic status, and human influence play crucial roles in moderating these effects. The immediate reductions in deforestation rates previously highlighted by the paired t-test suggest that targeted mitigation strategies can be effective. However, the broader impact of soyabean facility as shown by logistic regression, underscores the need for comprehensive, long-term strategies that incorporate both immediate and pervasive influences to effectively manage and reduce deforestation. The paired t-test results indicated a reduction in deforestation rates. The paired t-test measures the change in deforestation rates before and after the establishment of soyabean facilities, capturing the immediate impact and effectiveness of mitigation strategies. In contrast, the computed odds in this case of deforestation assesses the long-term effects of soyabean facilities on deforestation by calculating how likely deforestation is to occur in areas with these facilities (Hoffman, 2019).

4.5. Study implications and limitations.

4.5.1. Limitation

This study was not without limitations and uncertainties. The Mapbiomas maps though detailed only have an accuracy of 88.1%, which means that there was misclassification and this affects the quality of the analysis done on the images such as reclassifications. Additionally, errors in interpolation may arise from resampling the variables such as topological and soil factors as well as from transforming vector socioeconomic data to raster format. Errors from manual inspection and the limits of historical imagery also affect validation using Google Earth Pro. Furthermore, propensity score matching relies on the assumption that all relevant confounders are observed and included in the model known as the assumption of no unmeasured confounders or strong ignorability. If unobserved confounders affect both the treatment and the outcome, the matching process may not eliminate bias leading to incorrect causal inferences (Stuart, 2010). This assumption heavily depends on the quality and comprehensiveness of the data collected. All relevant variables that could potentially confound the relationship between the treatment and the outcome must be included in the propensity score model. However, not all variables may not have been included in our analysis such as temperature and policy variables.

These elements emphasize the necessity for cautious data collection and processing, validation, and cross-verification with other sources as well as model specifications since they can result in misclassification, spatial errors, temporal misalignment and biases in the study.

4.5.2. Implications

Regarding the social implications of the study, the land cover analysis provided a comprehensive understanding of how land cover has changed in Mato Grosso from 2002 to 2017. This highlighted the extent to which forests have been converted to agricultural lands (i.e., soyabeans, pastures, and mosaic agriculture) emphasizing the significant impact of agricultural expansion on deforestation. Additionally, the analysis of deforestation rates over time within different buffer zones offers insights into the spatial and temporal dynamics of deforestation. Understanding these patterns helped in identifying critical periods and areas where deforestation rates were highest and where mitigation efforts were most effective.

The study is crucial for government environment policymakers and environmental managers as it indicates that agricultural infrastructure development directly influences land use patterns and deforestation rates. Furthermore, the observed reduction in deforestation rates particularly in areas close to soyabean facilities suggests that certain mitigation strategies such as improved monitoring and sustainable agricultural practices (i.e., crop intensification) are effective. This supports the need for continued and enhanced policy measures targeting high-impact areas. The study's analysis of variables such as GDP per capita, travel time to the nearest city and human influence index provides valuable insights into how socioeconomic factors influence deforestation. This highlights the importance of integrating economic development and conservation efforts. Findings related to the influence of aspect and slope on deforestation provide a nuanced understanding of how physical geography affects land use decisions. This can inform land management practices and the designation of protected areas.

The study findings also provide insight into the benefits of adopting sustainable practices to agricultural producers. Producers may face stricter regulations and need to adapt to sustainable practices which could initially increase operational costs but ultimately lead to more sustainable and profitable farming methods. The owners of soyabean grain storage facilities benefit from the study findings as it provides a broader understanding of the impacts of their infrastructure on deforestation and land use patterns. This awareness can help them align their operations with sustainable practices, potentially improving their market reputation and compliance with environmental regulations. Owners may face increased pressure to adopt and implement sustainable practices to mitigate deforestation. This could involve investing in more sustainable infrastructure and technologies potentially increasing short-term costs but leading to long-term operational efficiencies and positive environmental impact. By adhering to sustainable practices, they can also avoid penalties and gain favour with environmentally conscious consumers and regulatory bodies.

Lastly, the study adds valuable data to the body of knowledge on land use changes and deforestation dynamics offering a foundation for further research on the impact of agricultural expansion. Researchers can build on the findings to explore additional factors influencing deforestation, develop new methodologies for land cover analysis and assess the long-term effectiveness of policy interventions.

Chapter Five: Conclusion and recommendation

5.1. Conclusion

This study has investigated to what degree soyabean grain storage facilities contribute to deforestation in Brazil using 5km, 10km and 25km buffers as observational framework. Essentially, these defined areas were used to systematically observe and quantify deforestation patterns in relation to the proximity to soyabean storage facilities. In addition, the buffers provided a framework in which the odds of deforestation were assessed with respect to the treatment, biophysical, socioeconomic and climates factors.

The study explored how land cover has changed by assessing deforestation rates within the different buffer zones from 2002 to 2017. Notable decline in deforestation rates from 2003 to 2008 are observed across buffers followed by stabilization at lower levels. This trend is particularly interesting as it aligns with observed trends in deforestation rate in Mato Grosso and reflects the effectiveness of certain policies and measures implemented during this period such as the soy moratorium and the Action Plan for Deforestation Prevention and Control in the Legal Amazon (PPCDAm). The initial high rates of deforestation in the 5 km buffer emphasize the immediate impact of soyabean storage facilities on nearby forest areas providing a clear temporal and spatial perspective on deforestation dynamics.

The study also explored the spatial patterns in deforestation and land cover changes. The paired t-test results indicate that the establishment of these facilities is associated with a substantial reduction in deforestation rates within defined buffer zones particularly within the 5 km buffer. This reduction is likely due to enhanced monitoring, enforcement, and sustainable practices implemented by agricultural producers which are more effective in closer proximities. Although the reduction in deforestation rates diminishes slightly within the 10 km and 25 km buffers, the impact remains significant highlighting the broader influence of sustainable land use practices and policy enforcement.

The analysis of specific land cover changes around soyabean storage facilities demonstrates a marked transformation in land use patterns, especially within the 5 km buffer, where a significant reduction in pasture areas suggests a repurposing of land previously used for grazing. This trend extends to the 10 km and 25 km buffers, indicating that the influence of soyabean storage facilities promotes sustainable practices and balanced land use over larger areas. The presence of protected areas and effective policy enforcement further supports the reduction in deforestation rates and the adoption of sustainable agricultural practices.

Regarding the socio-economic, biophysical factors on deforestation, the study observed that slope significantly reduces the likelihood of deforestation emphasizing the importance of geographical features in land use planning and conservation efforts. Higher GDP per capita is associated with reduced deforestation rates, highlighting the need for economic development to align with environmental sustainability initiatives. Furthermore, areas with higher human management and oversight show a slight reduction in deforestation odds underscoring the positive role of community engagement in conservation efforts.

The findings underscore the effectiveness of policies such as the soy moratorium and the CAR Registry in reducing deforestation. The study's results align with broader national and international trends observed by other research and monitoring projects, reinforcing the importance of sustained policy enforcement and innovation. This alignment validates the approaches taken by policymakers and provides a strong case for continuing and expanding these efforts.

5.2. Recommendation

Based on the findings of this study, the following recommendations are made to improve application of the research:

1. Inclusion of other data sources: Endeavour to make use of higher accuracy and spatial resolution datasets to minimize errors and misclassification. Also include additional relevant variables such as temperature and policy impacts variables in future analyses to enhance the robustness of the findings.
2. Enhanced monitoring and enforcement: Continue and expand monitoring and enforcement efforts near agricultural facilities to ensure compliance with sustainable practices. Implement more stringent oversight within immediate vicinities of new agricultural developments.
3. Socioeconomic development: Integrate economic development with conservation efforts. Increase funding and resources for conservation in areas with higher GDP per capita and promote economic activities that do not incentivize deforestation.
4. Geographical considerations in land management: Incorporate geographic features such as slope and aspect into land management and planning. Designate as protected areas regions with challenging terrains to leverage natural barriers against deforestation.
5. Generalizability: While the study focuses on Mato Grosso, the findings have broader implications for other regions experiencing similar agricultural expansion and deforestation dynamics. Comparative studies can build on these results to enhance global understanding of the impact of agricultural storage infrastructures on deforestation.

List of references

Álvaro, N. de S., Humberto, Â., John, E. B., Maísa, S. J., & Sandro, N. de S. (2012). Economic Feasibility of an Eucalyptus Agroforestry System in Brazil. INTECH Open Access Publisher.

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharmaceutical Statistics, 10(2), 150–161. https://doi.org/10.1002/pst.433

Bellón, B., Blanco, J., Vos, A. De, Roque, F. de O., Pays, O., & Renaud, P. C. (2020). Integrated Landscape change analysis of protected areas and their surrounding landscapes: Application in the Brazilian cerrado. Remote Sensing, 12(9). https://doi.org/10.3390/RS12091413

Brandão, M. V., Garbino, G. S. T., Semedo, T. B. F., Feijó, A., Do Nascimento, F. O., Fernandes-Ferreira, H., Rossi, R. V., Dalponte, J., & Carmignotto, A. P. (2019). Mammals of Mato Grosso, Brazil: Annotated species list and historical review. Mastozoologia Neotropical, 26(2), 263–307. https://doi.org/10.31687/saremMN.19.26.2.0.03

Busch, J., & Ferretti-Gallon, K. (2017). What drives deforestation and what stops it? A meta-analysis. Review of Environmental Economics and Policy, 11(1), 3–23. https://doi.org/10.1093/reep/rew013

De Maria, M., Robinson, E. J., Kangile, J. R., Kadigi, R., Dreoni # °, I., Couto, M., Howai, N., Peci, J., & Fiennes, S. (2020). Global Soyabean Trade-The Geopolitics of a Bean. UK Research and Innovation Global Challenges Research Fund (UKRI GCRF) Trade, Development and the Environment Hub. https://doi.org/10.34892/7yn1-k494

DeFries, R., Herold, M., Verchot, L., Macedo, M. N., & Shimabukuro, Y. (2013). Export-oriented deforestation in Mato Grosso: Harbinger or exception for other tropical forests? Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1619). https://doi.org/10.1098/rstb.2012.0173

Del, M., Vera-Diaz, C., Kaufmann, R. K., & Nepstad, D. C. (2009). The Environmental Impacts of Soyabean Expansion And Infrastructure Development in Brazil's Amazon Basin. http://ase.tufts.edu/gdae

Eberhardt, I. D. R., Teixeira, L. M. A., Granemann, S. R., Fonseca, A. P., Taco, P. W. G., & Furtado, B. A. (2017). Supply-demand of grain storage facilities in Mato Grosso state in Brazil. Proceedings - 2017 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI 2017, 2017-January, 58–62. https://doi.org/10.1109/SOLI.2017.8120970

Ebregt, A., & De Greve, P. (2000). Buffer zones and their management Policy and Best Practices for terrestrial ecosystems in developing countries Theme Studies Series 5 Forests, Forestry and Biological Diversity Support Group. www.minlnv.nl

Erasmus, K. H. J. zu E., Ben, A., Javier, G., Mairon, G. B. L. , Simone, B., Rachael, G., Jonathan, G., Michael J Lathuillière, Pernilla, L., Christina, M., Patrick, M., Clément, S., Chris, W., & Toby, G. (2020). Using supply chain data to monitor zero deforestation commitments: an assessment of progress in the Brazilian soy sector. https://doi.org/10.1088/1748-9326/ab6497

Fehlenberg, V., Baumann, M., Gasparri, N. I., Piquer-Rodriguez, M., Gavier-Pizarro, G., & Kuemmerle, T. (2017). The role of soyabean production as an underlying driver of deforestation in the South American Chaco. https://doi.org/10.1016/j.gloenvcha.2017.05.001

Fernando, C. de O., Paulo, H. M. C., Maurílio, de S. N., Andre, C. de S. A., Fenelon, L. de S. S., Joao, P. de M. O., Bruna, S. de O., Itamar, R. T., & Andre, J. de C. (2016). Logistics and storage of soyabean in Brazil. African Journal of Agricultural Research, 11(35), 3261–3272. https://doi.org/10.5897/ajar2016.11198

Filassi, M., & de Oliveira, A. L. R. (2021). Competitiveness drivers for soyabean exportation and the fundamental role of the supply chain. Revista de Economia e Sociologia Rural, 60(3), 1–22. https://doi.org/10.1590/1806-9479.2021.235296

Garrett, R. D., Lambin, E. F., & Naylor, R. L. (2013). Land institutions and supply chain configurations as determinants of soyabean planted area and yields in Brazil. Land Use Policy, 31, 385–396. https://doi.org/10.1016/j.landusepol.2012.08.002

Garrido, M. M., Kelley, A. S., Paris, J., Roza, K., Meier, D. E., Morrison, R. S., & Aldridge, M. D. (2014). Methods for constructing and assessing propensity scores. Health Services Research, 49(5), 1701–1720. https://doi.org/10.1111/1475-6773.12182

Gladek, E., Fraser, M., Roemers, G., Sabag Muñoz, O., Kennedy, E., Hirsch, P., Bjorck, C., Baheranwala, F., Lakey, M., Manisha, P., Donald, M., Jameson, S., Bouwens, J., Sontag, H., Marselis, I., Keatinge, T., Speyer, E., Huang, E., Doornbos, S., … Damian, O. (2017). The global food system:an analysis.

Graesser, J., & Ramankutty, N. (2017). Detection of cropland field parcels from Landsat imagery. Remote Sensing of Environment, 201, 165–180. https://doi.org/10.1016/j.rse.2017.08.027

Hoffman, J. I. E. (2019). Logistic Regression. Basic Biostatistics for Medical and Biomedical Practitioners, 581–589. https://doi.org/10.1016/B978-0-12-817084-7.00033-4

Krejcie, R. V, & Morgan, D. W. (1970). Determining sample size for research activities. in educational and psychological measurement (Vol. 30).

Kuschnig, N., Cuaresma, J. C., Krisztin, T., & Giljum, S. (2021). Spatial spillover effects from agriculture drive deforestation in Mato Grosso, Brazil. Scientific Reports, 11(1). https://doi.org/10.1038/s41598-021-00861-y

López-Carr, D. (2021). A review of small farmer land use and deforestation in tropical forest frontiers: Implications for conservation and sustainable livelihoods. Land, 10(11). https://doi.org/10.3390/land10111113

Macedo, M. N., DeFries, R. S., Morton, D. C., Stickler, C. M., Galford, G. L., & Shimabukuro, Y. E. (2012). Decoupling of deforestation and soy production in the southern Amazon during the late 2000s. Proceedings of the National Academy of Sciences of the United States of America, 109(4), 1341–1346. https://doi.org/10.1073/pnas.1111374109

Masuda, T., & Goldsmith, P. D. (2009). World Soyabean Production: Area Harvested, Yield, and Long-Term Projections. International Food and Agribusiness Management Review, 12(4), 1–20. https://doi.org/10.22004/AG.ECON.92573

Meier, R., Schwaab, J., Seneviratne, S. I., Sprenger, M., Lewis, E., & Davin, E. L. (2021). Empirical estimate of forestation-induced precipitation changes in Europe. Nature Geoscience, 14(7), 473–478. https://doi.org/10.1038/s41561-021-00773-6

Morton, D. C., DeFries, R. S., Shimabukuro, Y. E., Anderson, L. O., Arai, E., del Bon Espirito-Santo, F., Freitas, R., & Morisette, J. (2006). Cropland expansion changes deforestation dynamics in the southern Brazilian Amazon. www.obt.inpe.brdeter.

Müller, R., Müller, D., Schierhorn, F., & Gerold, G. (2011a). Spatiotemporal modeling of the expansion of mechanized agriculture in the Bolivian lowland forests. Applied Geography, 31(2), 631–640. https://doi.org/10.1016/j.apgeog.2010.11.018

Nguyen, T. L., Collins, G. S., Spence, J., Daurès, J. P., Devereaux, P. J., Landais, P., & Le Manach, Y. (2017). Double-adjustment in propensity score matching analysis: choosing a threshold for considering residual imbalance. BMC Medical Research Methodology, 17(1). https://doi.org/10.1186/s12874-017-0338-0

Perelló, L. F. C., Guadagnin, D. L., Maltchik, L., & dos Santos, J. E. (2012). Ecological, legal, and methodological principles for planning buffer zones. Natureza a Conservacao, 10(1), 3–11. https://doi.org/10.4322/natcon.2012.002

Rosa, I. M. D., Purves, D., Souza, C., & Ewers, R. M. (2013). Predictive Modelling of Contagious Deforestation in the Brazilian Amazon. PLoS ONE, 8(10). https://doi.org/10.1371/journal.pone.0077231

Ross, A., & Willson, V. L. (2017). Paired Samples T-Test. Basic and Advanced Statistical Tests, 17–19. https://doi.org/10.1007/978-94-6351-086-8_4

Roy, P. S., Ramachandran, R. M., Paul, O., Thakur, P. K., Ravan, S., Behera, M. D., Sarangi, C., & Kanawade, V. P. (2022). Anthropogenic Land Use and Land Cover Changes—A Review on Its Environmental Consequences and Climate Change. In Journal of the Indian Society of Remote Sensing (Vol. 50, Issue 8, pp. 1615–1640). Springer. https://doi.org/10.1007/s12524-022-01569-w

Shaik, S., & Kwame-Asiam, A. (2020). Soyabean Production Indicators Report: Trend and Risk Analysis.

Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. American Journal of Applied Mathematics and Statistics, 8(2), 39–42. https://doi.org/10.12691/ajams-8-2-1

Soares-Filho, B., & Rajão, R. (2018). Traditional conservation strategies still the best option. In Nature Sustainability (Vol. 1, Issue 11, pp. 608–610). Nature Publishing Group. https://doi.org/10.1038/s41893-018-0179-9

Song, X., Huang, W., Hansen, M. C., & Potapov, P. (2021). An evaluation of Landsat, Sentinel-2, Sentinel-1 and MODIS data for crop type mapping. Science of Remote Sensing, 3, 100018. https://doi.org/10.1016/j.srs.2021.100018

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. In Statistical Science (Vol. 25, Issue 1, pp. 1–21). https://doi.org/10.1214/09-STS313

Szerman, D., Assunção, J. J. ;, Lipscomb, M., & Mobarak, A. M. (2022). Standard-Nutzungsbedingungen. https://hdl.handle.net/10419/265572

The Sustainable Development Goals. (2015).

Vieira, D. C., Sanches, I. D. A., Montibeller, B., Prudente, V. H. R., Hansen, M. C., Baggett, A., & Adami, M. (2022). Cropland expansion, intensification, and reduction in Mato Grosso state, Brazil, between the crop years 2000/01 to 2017/18. Remote Sensing Applications: Society and Environment, 28. https://doi.org/10.1016/j.rsase.2022.100841

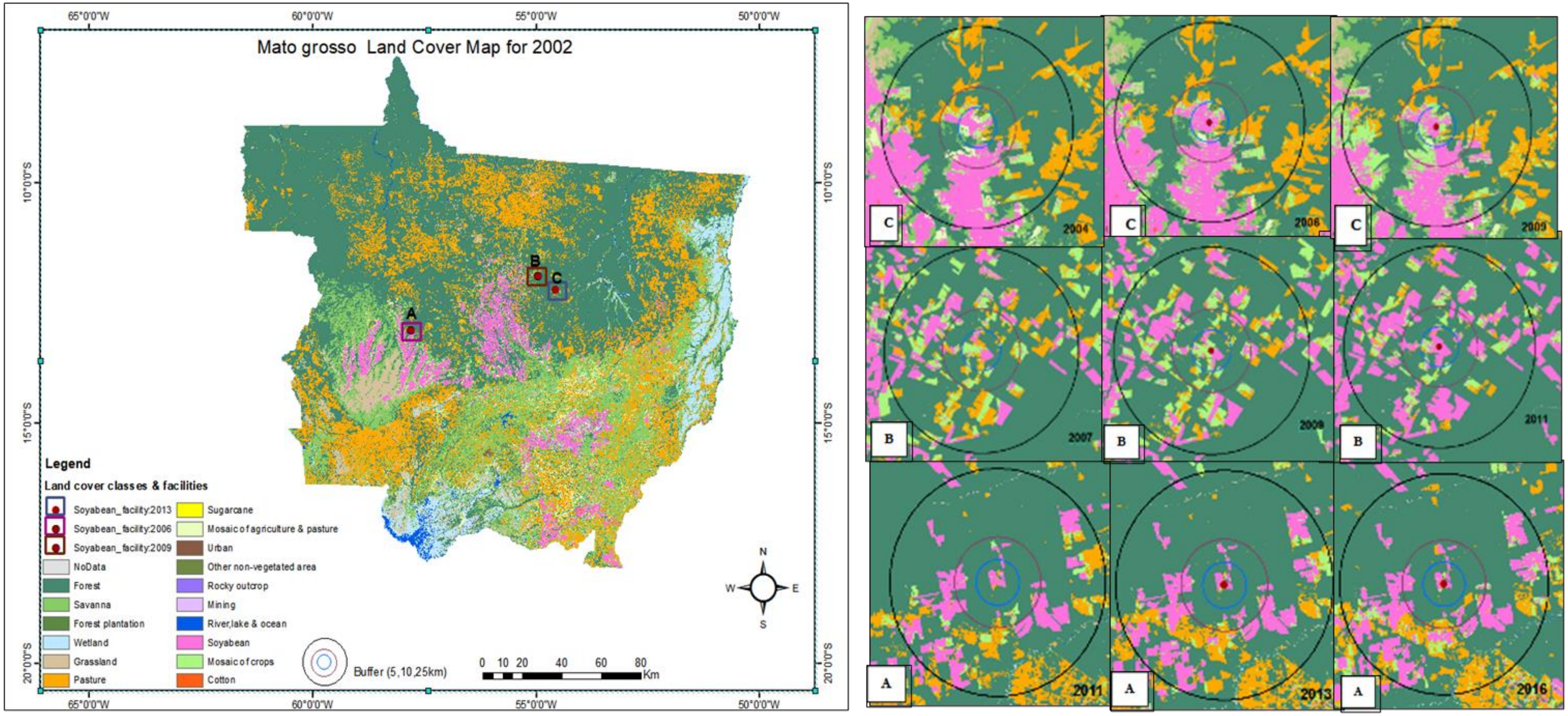
Vivek, V., Cristina, L., & Steffany, B. (2020). Global Market Report: Soyabeans. https://www.jstor.org/stable/resrep26554

Winkler, K., Fuchs, R., Rounsevell, M., & Herold, M. (2021). Global land use changes are four times greater than previously estimated. Nature Communications, 12(1). https://doi.org/10.1038/s41467-021-22702-2

Yang, H., Viña, A., Winkler, J. A., Chung, M. G., Huang, Q., Dou, Y., McShea, W. J., Songer, M., Zhang, J., & Liu, J. (2021). A global assessment of the impact of individual protected areas on preventing forest loss. Science of the Total Environment, 777. https://doi.org/10.1016/j.scitotenv.2021.145995

Appendix

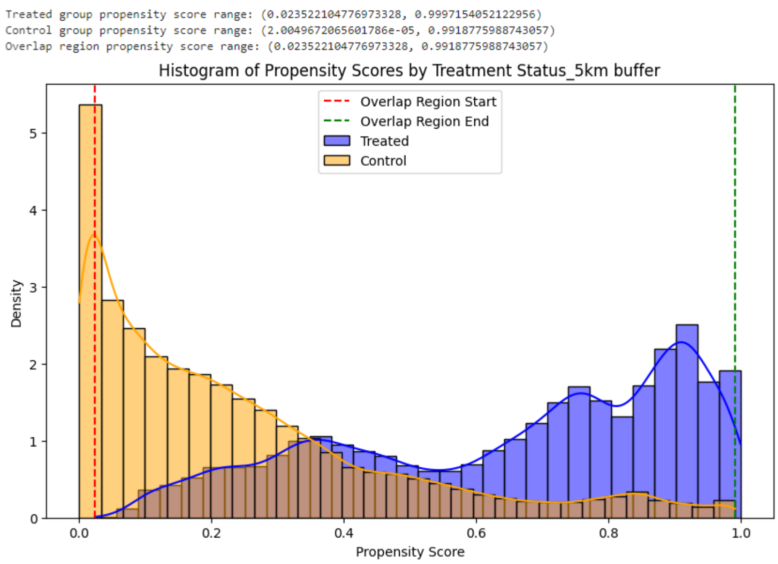
Appendix 1: Land cover changes within soyabean grain facilities built in 2006(C), 2009 (B) and 2013 (A).

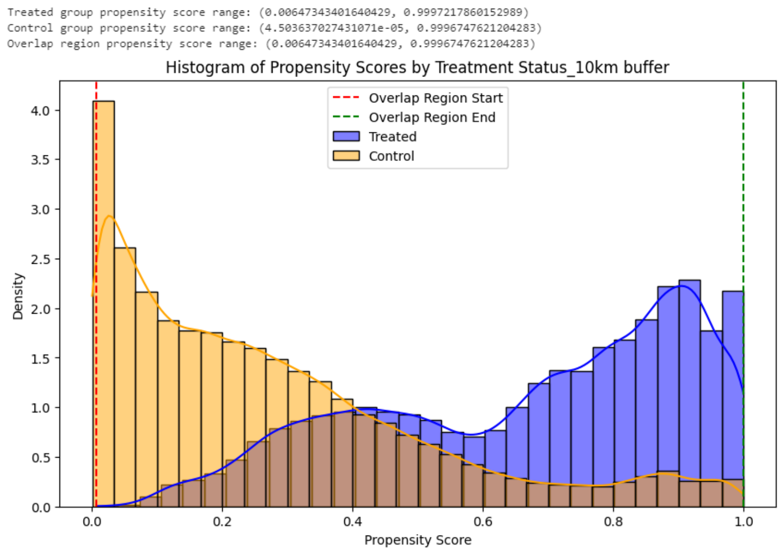


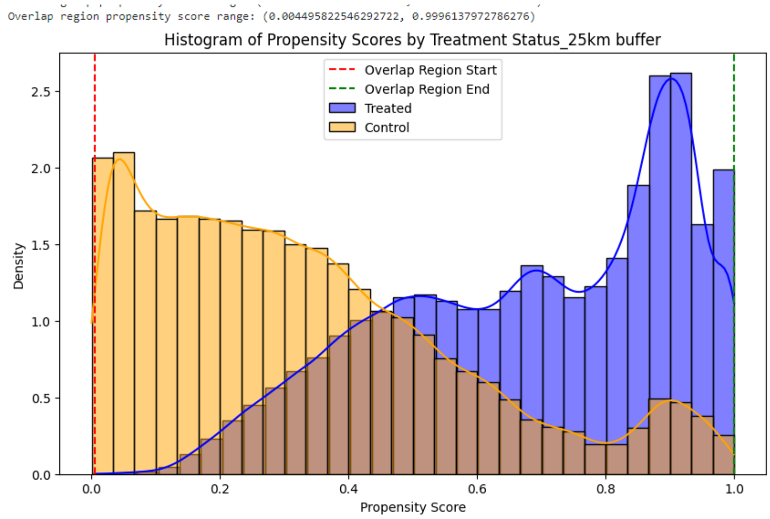
Appendix 2: Matched and unmatched units by buffer zone

Buffer Zone	Matched Treated Units	Matched Control Units	Unmatched Treated Units	Unmatched Control Units
5 km	109,081	109,081	0	121,521
10 km	136,876	136,876	0	118,998
25 km	173,270	173,270	0	108,375

Appendix 3: Propensity score distribution matching output

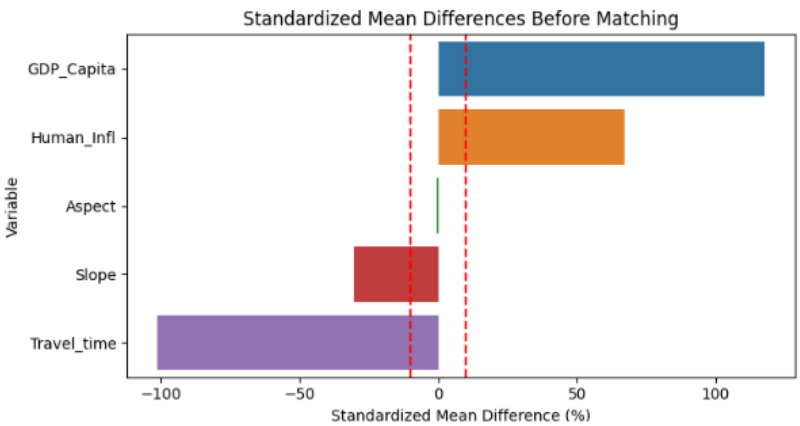


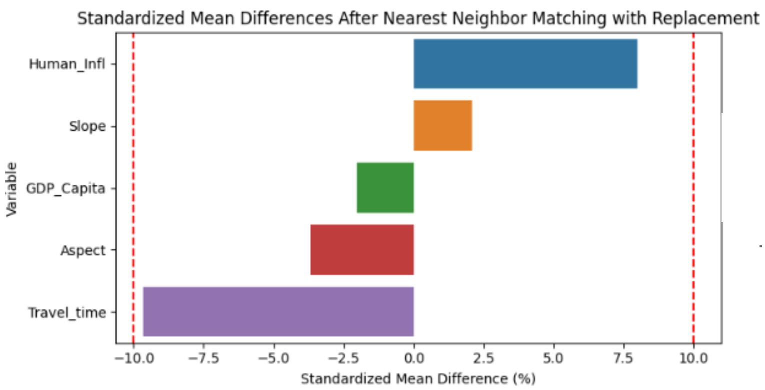




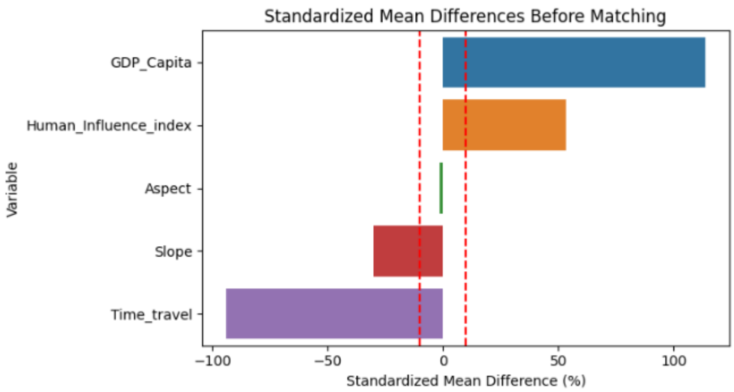
Appendix 4: Standard mean difference

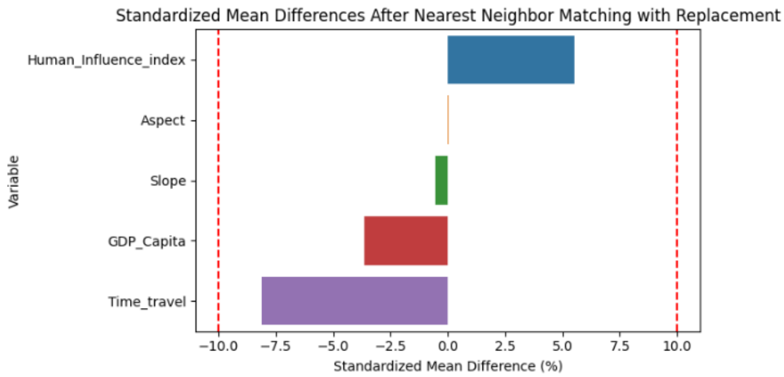
5km buffer



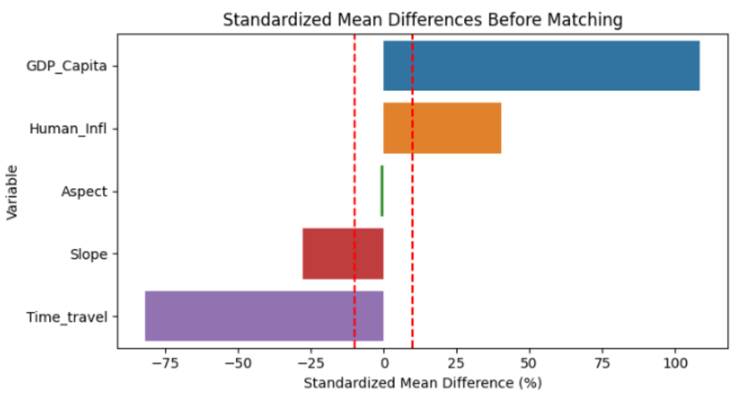


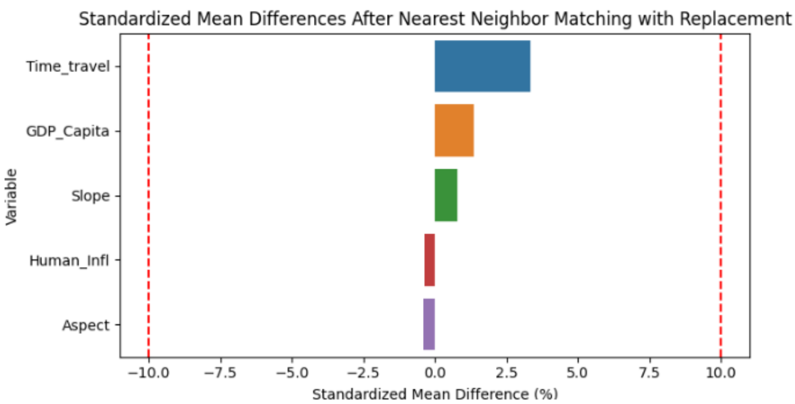
10km buffer





25km buffer





Appendix 5: Odd ratio estimates

Variable	Odds Ratio	Std.err.	z	p>|z|	95% conf. interval	
					Lower	Upper
Const.	0.352654	0.022149	-47.056922	0.000	0.337672	0.368300
Soyabean grain facility	2.199309	0.011004	71.623845	0.000	2.152383	2.247257
Aspect	1.000176	0.000050	3.502117	0.000	1.000077	1.000274
Slope	0.849650	0.004045	-40.275549	0.000	0.842939	0.856413
GDP_Capita	0.984699	0.000331	-46.543446	0.000	0.984060	0.985339
Travel_time	1.000628	0.000018	34.825982	0.000	1.000592	1.000663
Human influence index	0.986764	0.000861	-15.481575	0.000	0.985101	0.988430

10km buffer

Variable	Odds Ratio	Std.err.	z	p>|z|	95% conf. interval	
					Lower	Upper
Const.	0.371740	0.019941	-49.624669	0.000	0.357491	0.386557
Soyabean grain facility	1.768814	0.009935	57.401872	0.000	1.734703	1.803595
Aspect	1.000305	0.000046	6.648140	0.000	1.000215	1.000395
Slope	0.843845	0.003777	-44.948607	0.000	0.837621	0.850116
GDP_Capita	0.984809	0.000290	52.757923	0.000	0.984249	0.985369
Travel_time	1.000466	0.000016	28.215039	0.000	1.000433	1.000498
Human influence index	0.988400	0.000869	-13.420809	0.000	0.986717	0.990086

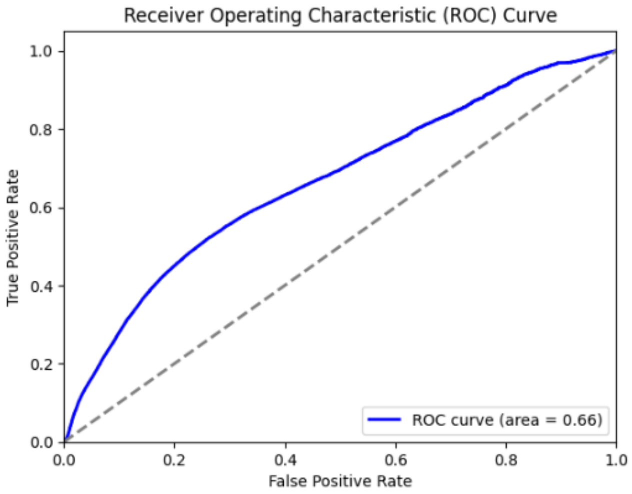
25km buffer

Variable	Odds Ratio	Std.err.	z	p>|z|	95% conf. interval	
					Lower	Upper
Const.	0.430606	0.018111	-46.522892	0.000	0.415590	0.446166
Soyabean grain facility	1.471436	0.008912	43.337585	0.000	1.445956	1.497364
Aspect	1.000416	0.000041	10.094449	0.000	1.000335	1.000496
Slope	0.844263	0.003404	-49.732158	0.000	0.838649	0.849915
GDP_Capita	0.990965	0.000246	-36.924785	0.000	0.990488	0.991442
Travel_time	0.999946	0.000015	-3.601390	0.000	0.999917	0.999976
Human influence index	0.969168	0.000889	-35.231780	0.000	0.967481	0.970858

Appendix 6: Measures of model fitness

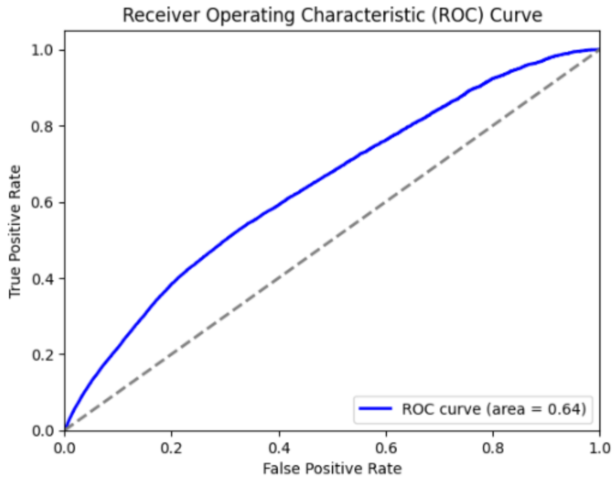
5km buffer

AIC: 214845.06381881674
ROC-AUC: 0.6636512242418487



10km buffer

AIC: 260956.52691747888
ROC-AUC: 0.6401505951452902



25km buffer

AIC: 323275.12379112653
ROC-AUC: 0.6167982885174759